AD________________

Award Number: W81XWH-06-2-0057

TITLE: Operating Room of the Future: Advanced Technologies in Safe and Efficient Operating Rooms

PRINCIPAL INVESTIGATOR: Adrian E. Park, M.D.

CONTRACTING ORGANIZATION: University of Maryland Medical Center
Baltimore, MD 21201

REPORT DATE: October 2008

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE
1 Oct 2008

2. REPORT TYPE
Annual

3. DATES COVERED
26 Sep 2006 – 25 Sep 2008

4. TITLE AND SUBTITLE

Operating Room of the Future: Advanced Technologies in Safe and Efficient Operating Rooms

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-2-0057

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Adrian E. Park, M.D.

E-Mail: gmoses@smail.umaryland.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

University of Maryland Medical Center
Baltimore, MD 21201

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The current research project activities are based upon three pillars of research, OR Informatics, Simulation for Training and Smart Image. A fourth pillar was added during this period of performance that targeted physical and cognitive ergonomics/human factors. At the beginning of this period of performance, there were five projects that comprised the Informatics pillar and two for Smart Image. The Simulation pillar had been comprised of a single project, The Maryland Virtual Patient. This pillar was expanded to include research conducted in and for the larger Simulation Training Program in the MASTRI Center. Finally, work objectives were established for research related to physical and cognitive ergonomics/human factors. This research has proceeded under the mantle of the phrase "Operating Room of the Future'. We are replacing that theme with the more appropriate "Innovations in the Surgical Environment". The period of performance of this contract was extended to October 30, 2009. Based on the extended period of performance, this project is on time, on schedule, and within performance parameters.

15. SUBJECT TERMS
Advanced medical technology, Surgical innovation, informatics, simulation, smart image, human factors.

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** USAMRMC
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	176	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

2. Current staff, role and percentage effort on each project.

Staff Member	Role
A. Park, MD	PI
B. Jarrell, MD	Co-PI
G. Moses, PhD	Director
S. Kavic, MD	Clinical Support
P. Turner, MD	Clinical Support
G. Fantry, MD	Clinical Support
R. Shekhar, PhD	Smart Image
P. Nagy, PhD	Informatics Research
Y. Xiao, PhD	Informatics

3. Contract Expenditures to date (as applicable)

Cost Elements	Cumulative
Personnel	833,231
Fringe Benefits	134,931
Supplies	53,018
Equipment	1,900
Travel	11,690
Other direct costs	642,032
Subtotal	1,676,802
Indirect Costs	304,037
Fee	
Total	1,980,839

Introduction

During the past decade, we witnessed an extraordinary evolution in surgical care based upon rapid advances in technology and creative approaches to medicine. The increased speed and power of computer applications, the rise of visualization technologies related to imaging and image guidance, improvement in simulation-based technologies (tissue properties, tool-tissue interaction, graphics, haptics, etc) has caused an explosion in surgical advances. That said, we remain far behind scientists in applying information systems to patient care. This research effort has proceeded under the mantle of "Operating Room of the Future" research. We are replacing that theme with the more appropriate "Innovations in the Surgical Environment."

The content of this annual report contains information pertinent to continued activities in relation to the W81XWH-06-2-005, "Advanced Technologies in Safe and Efficient Operating Rooms" project. This contract consists of a scope of work that fits seamlessly onto a prior research activity in the contract DAMD-17-03-2-0001, "Advanced technologies in safe and efficient operating rooms" work. The current research project activities are based upon three pillars of research, OR Informatics, Simulation for Training and Smart Image. A fourth pillar was added during this period of performance that targeted physical and cognitive ergonomics/human factors.

At the beginning of this period of performance, there were five projects that comprised the Informatics pillar and two for Smart Image. The Simulation pillar had been comprised of a single project, The Maryland Virtual Patient. This pillar was expanded to include research conducted in and for the larger Simulation Training Program in the MASTRI Center. Finally, work objectives were established for research related to physical and cognitive ergonomics/human factors.

We recently conducted our annual conference, Innovations in the Surgical Environment, a meeting that serves as a deep recapitulation of research performed under the contract. Additionally, the conference presents an opportunity to explore innovative approaches to surgical research with government, academic and industry partners, and expands our capability to develop collaborative relationships. This year, the conference theme was lessons learned from the high-stakes environments of aviation and astronautics applied to the high-stakes environment of the operating room.

Body

A. OR Informatics

Informatics subgroup 1. Workflow and Operations Research for Quality (WORQ)

The Perioperative Scheduling Study examined how using post-operative destination information during the process of surgery scheduling can influence congestion in post-operative units such as ICUs and IMCs, which lead to overnight boarders in the PACU. The research team is Jeffrey W. Herrmann, Ph.D., and Greg Brown, a graduate student, both with the University of Maryland, College Park. The team works closely with Michael Harrington, Ramon Konewko, R.N., and Paul Nagy, Ph.D., for guidance and assistance.

The surgery scheduling process has been carefully studied to understand the different organizations and persons who participate in the process, include the schedulers in the surgical services, the perioperative services office, the PACU manager, and the OR charge nurse. Interviews with many of these groups and observations of their scheduling process were conducted on January 17, 2008. These groups also provided copies of their scheduling policies and typical schedules.

We have developed a mathematical model for evaluating congestion in post-operative units, including ICUs, IMCs, and floor units. This model requires data about post-operative destinations and length-of-stay distributions for different types of surgeries. We have analyzed data about cardiac surgeries from two years and have analyzed UMMC financial records for all of the surgical cases for fiscal year 2007. We have developed an algorithm for predicting bed requirements based on the surgical schedule and have conducted a preliminary study comparing these predictions to other prediction methods for two units. The preliminary results show that the new bed requirements prediction method is more accurate. We plan to complete the study and document the results in a technical report this summer. We have also begun considering approaches for predicting bed requirements when the surgical schedule is incomplete.

Major Accomplishments achieved during this period of performance include the successful analysis of patient data for all surgical cases, the development of an algorithm for predicting bed requirements, and the comparison of a variety of prediction methods for two units.

Informatics subgroup 2. Operating Room Glitch Analysis (OGA)

The OGA project, focusing on institutional learning, examined the workflow around performance indicators in the perioperative environment and building a graphical dashboard to allow data mining and trend analysis of operating indicators.

The dashboard was constructed using the Ruby on Rails web development platform with a MySQL database dynamically driving the queries. An interactive graphical dashboard provided synthesis around delays in operations with multiple information visualization techniques.

The initial surgical dashboard provided a strategic view of the department. In order to sustain the current system and make it flexible to the growing need for information several infrastructure projects were undertaken. The results are as follows:

- A data warehouse was created to store information in real time or in polling intervals
- A user interface was created for analysts to perform data manipulation and define new metrics without a knowledge of programming
- The dashboard was updated to allow users to define how data should be visualized
    - Users can create charts with a drag and drop interface
    - Charts can then be saved into a web page accessible by surgery personnel
- A dedicated server was obtained and setup to host both the warehouse and the dashboard

The data warehouse and user interface provide a supportable and scalable platform for the current surgical dashboard. Furthermore, these infrastructure pieces empower the users of the system to adjust and add new metrics as needed. As transactional information pipelines are built with vendor cooperation the dashboard will be able to incorporate tactical metrics to improve workflow and patient care in real time.

We have integrated into the data architecture a javascript based bubble chart that provides several interactive features to allow thorough data discovery. The bubble chart can play over time to see how the size of the bubbles change, which relates to the number of cases performed, as well as their x and y axis location. The x and y axis can represent delay duration, actual procedure time, scheduled procedure time, or turnover time. The bubble can also be tagged to provide a contrail to show performance over time.

The dashboard system is described in detail in an article published in Surg Innov 2008 15: 7-16, by Paul G. Nagy, Ramon Konewko, Max Warnock, Wendy Bernstein, Jacob Seagull, Yan Xiao, Ivan George, and Adrian Park, entitled “Novel, Web-Based, Information-Exploration Approach for Improving Operating Room Logistics and System Processes.

Informatics subgroup 3. Context Aware Surgical Training (CAST)

We proposed to design and implement a prototype context aware surgical training environment (CAST) as part of the University of Maryland Medical System's SimCenter. This system was designed to explore the role that an intelligent pervasive computing environment can play to enhance the training of surgical students, residents and specialists. The research built upon prior work on context aware "smart spaces" done at UMBC; leverage our experience in working with RFID in the DARPA Trauma Pod program as well as in incorporating Web-based infrastructure and software applications in academic and professional development programs. The project was expected to result in a pilot system integrating one or two training resources available in the SimCenter into a context aware training environment that can recognize the presence of a trainee and or mentor and take appropriate action based on known training goals and parameters. The project proposed to advance the knowledge of context aware training environments in a highly technical medical field and provide a basis for incorporating more advanced technology assisted learning experiences in medicine. This "smart environment" may then, if successful, be scaled to meet the needs of an operative environment where the technological demands may be the similar or analogous to those seen in the training environment. Ultimately, the advanced training and potential for use in perioperative environments have a long-term end goal of improving patient safety and adding to the body of knowledge in surgical training. Initially, we saw a situation were clinicians in training can receive a tailored curriculum. Additionally, we envisioned a system that offers real-time feedback and decision support and education metrics to faculty.

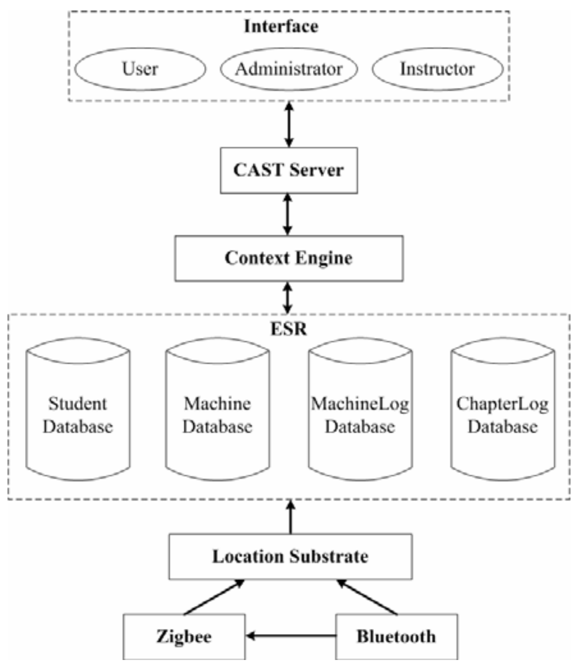
A key goal this year was to prototype the CAST system. To start off, we had meetings with the MASTRI team responsible for the training efforts led by Dr Turner to iron out the requirements for the system and came up with the following (initial) set of tasks to be accomplished

- Student Tracking
- Enforcing Prerequisites
- Video Capture
- Instructor Feedback

Also, we defined a typical use case for our system.
A Student enters the simulation center. The system identifies the student (for instance, using their Bluetooth phone or their badge), and does a prerequisite check based on the simulator the student wants to perform the procedure. Only if the student is done with the prerequisites, is he/she allowed to proceed. When the student indicates that they are ready to begin, the system starts capturing the external and internal view until the student indicates that they have completed the task. The captured video is then transferred to the video server for review by the instructor. The instructor interface allows the instructor to see the entry logs of students in terms of when they entered and exited the centre along with the corresponding external view.

We employed the spiral prototyping approach as an experimental test bed; we designed and implemented an initial system prototype that would meet the above functional requirements. The prototype integrates two machines with each simulator -- a small Nokia 800 device for resident interaction, and a larger PC for video capture. Note that this is for the proof of concept. A single small form factor but computationally powerful machine could be used instead. In fact, for simulators such as the VR, we expect that eventually manufacturers could integrate our system directly into the computer that drives the simulation.

Our prototype used Bluetooth for localization of residents in the simulation centre. It was designed to be modular, so that any other technology (such as resident ID cards) could be integrated easily. We also hosted training materials including videos for FLS, Kentucky and Rosser tasks in our system, and tracked student progress through the chapters checked out. This was used for enforcing prerequisites when students entered the simulation centre to perform procedures. In addition to enforcing prerequisites, there was a need for the instructors to visually see what the residents were doing during their simulation procedures. We use N800's built in camera to capture the residents' external views. These video feeds are then fed into a central server for review by the instructor.



For location detection, we also experimented with using the Awarepoint tags. Awarepoint uses a zigbee based mesh network for localization and exposes the location information through a web service. Our experiments indicated that Awarepoint could provide us room level information, but not anything finer. While this would help identify if the residents were in the simulation center, it would not help determine which machine they were using, which was needed for CAST. We demonstrated our first system prototype at the ORF workshop by going through a typical student workflow.

Based on feedback on individual components of the first prototype, we started the second version of the prototype to be deployed at the MASTRI center. The key changes in the second prototype from the first one are described below.

- We no longer use awarepoint for locationing since it could provide us only room level accuracy.
- On the student identification front, we are using a standard username and password method for now. Also, since we have external camera views from the N800, students identified can be verified during the review process by the instructor. Bluetooth based identification exists, but is not used since we were told that most residents may not have phones with Bluetooth. Multiple candidate technologies for identification such as Bluetooth, RFID, nearfield RF badges etc. have emerged and been discussed with MASTRI staff. No single choice has been made yet – the idea is to first make the system robust from a use perspective, and then integrate identification technologies based on further discussion with MASTRI staff. Our system is capable and flexible enough to handle a variety of lower level locationing technologies and therefore we would choose the one that is most practical in MASTRI scenario.
- Due to hospital network firewall policies, we had to move away from using a wireless network for transferring videos from the N800 to the MASTRI video server. We currently achieve this by tunneling through the internal view capturing machine which is hooked to the N800 by a USB cable.
- Prerequisite checks are temporarily suspended since the initial classes being taught in MASTRI are not following FLS.

We also focused on moving the system from UMBC machines to the MASTRI infrastructure where they will be housed. We purchased a small factor Dell machine to be used for capturing internal views from simulators. Storage was purchased and added to the mastri-internal server for archiving both internal and external video feeds with help from Jesus. Also, we have

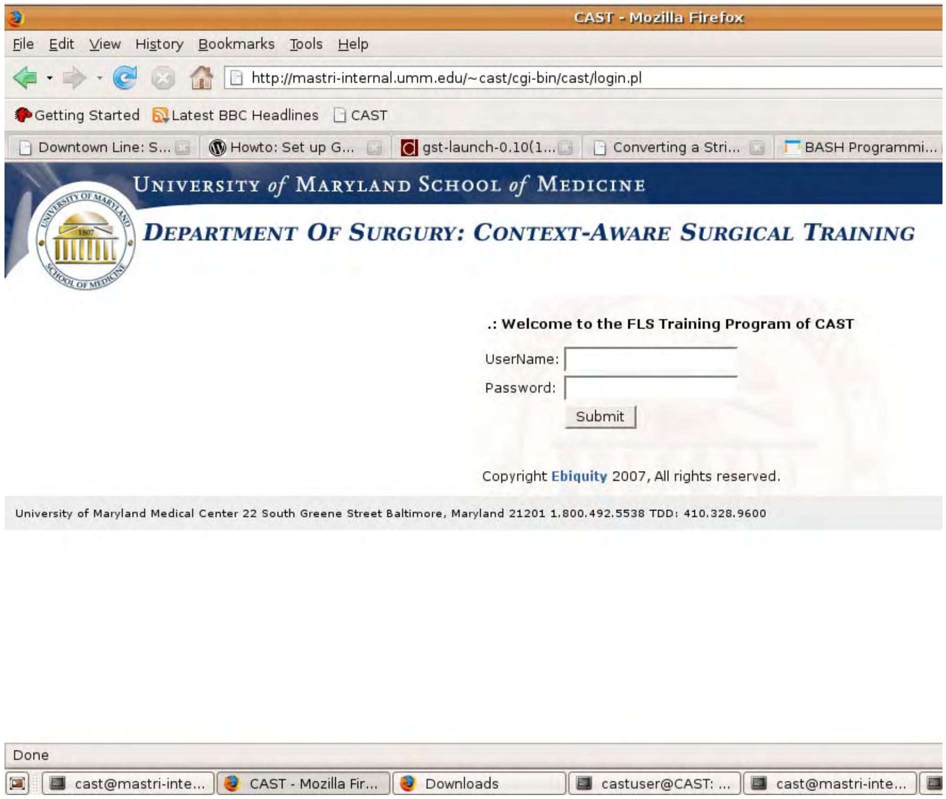
- Integrated the student database from the hospital
- Hosted FLS and other training videos on the hospital infrastructure
- Hacked internal views of the simulators

We developed the system to capture internal video feeds and metrics from the following simulators

- Promis
- Stryker
- Laproscopic VR simulator

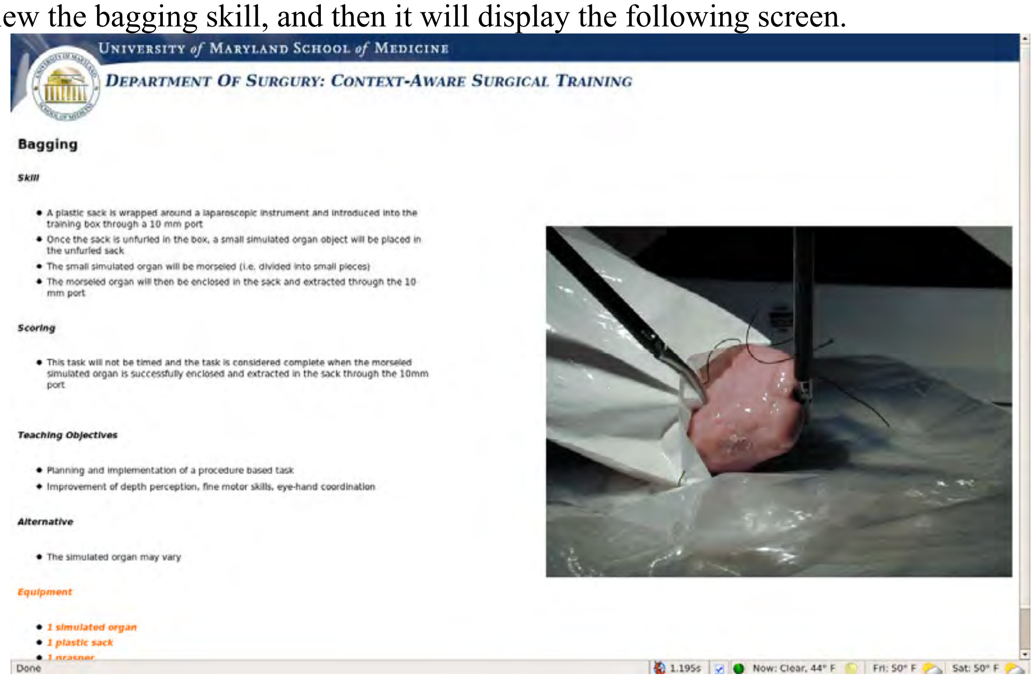
We use external s-video frame grabbers to capture the simulator internal video feeds. These feeds are synchronized with the external view from the N800 and stored on the video server. Thus, the instructor now has access to both the internal and external feeds during review, and consequently they can provide better feedback. Currently our system uses email to send back feedback.

Web Interface:



We have integrated the student database into our web-based curriculum management system. The student database contains all the current residents and one guest account that can be used for testing. If a student wants to view the training videos, he or she will first need to log in using their SMail Userid. When they pass the authentication, a categorized web structure will be displayed, and they can choose to sort the tasks by category, by difficulty (FLS integrated), or by each (Basic, Instrument, Procedural Skills, and FLS), which is shown in the following screenshot. This structure was developed in consultation with the MASTRI team, particularly Ivan George and Ethan Hagan.

Then the student can pick any training video they want to view. Suppose the student want to view the bagging skill, and then it will display the following screen.



As for the instructor interface, the instructor first needs to pass authentication to access the student training records. Then, they can pick the student name that they want to view from a drag-down list that contains all the residents. The appropriate student record will pop up in the next page, which contains the following information: the chapters that the student has checked out, the student training history (simulator type, start time, end time, internal video record, external video record), and the instructor can provide their feedback for every training record of this student via email.

Current efforts have focused on testing an initial deployment of the CAST system at the MASTRI Center. We demonstrated the system to a set of resident volunteers for feedback in a form of Beta-test of the system. We set up hardware and software to include the VR Simulator as part of the CAST system deployment. We got usability feedback and fixed bugs. A significant part of the effort was also spent in surveying the state-of-the-art in Video/VR usage for surgical training. We identified a small but significant body of work (e.g. Sinanan et al, Darzi et al) in checking the construct validity of the models for training using these simulation tools. The typical approach is to use sensors to capture the kinematics of the tools, as well as force/torque measures. The UMBC/MASTRI team decided that we would like to focus on an alternate approach that i) focused on the video, not (initially) any other sensors and ii) tried to capture using machine learning techniques the ability of an expert surgeon to identify key events in a surgery that relate to outcome or skill assessment. This is a very challenging and open problem. Key initial steps were identified for initial implementation in the first year.

A detailed description of the CAST project, “A Ubiquitous Context-Aware Environment for Surgical Training”, was presented at the First International Workshop on Mobile and Ubiquitous Context Aware Systems and Applications (MUBICA 2007), August 2007, by P. Ordóñez, P. Kodeswaran, V. Korolev, W. Li, O. Walavalkar, B. Elgamil, A. Joshi, T. Finin, Y. Yesha, I.George.

NOTE:

The effort to establish a system for archiving imaged data from training sites has been attenuated due to advancements in archiving capability in off-the-shelf systems. The focus of this project now rightly shifts to video summarization by unique application of artificial intelligence techniques. Video summarization has extraordinary potential for streamlining the events in the future perioperative environment. Further, there are many and varied military applications from video summarization. The UMBC Graduate students currently working on the CAST and the background research effort for the new direction will transition out, as the new direction is less closely aligned with their research interests. Dr. Mike Grasso, MD/PhD in Computer Science, will be joining the effort, and a new graduate student whose research focus will be on the video efforts will join the team. This is of course is subject to new funds being available.

Video Summarization.

The effort to establish a system for archiving imaged data from training sites has been attenuated due to advancements in archiving capability in off-the-shelf systems. The focus of this project now rightly shifts to video summarization by unique application of artificial intelligence techniques. Video summarization has extraordinary potential for streamlining the events in the future perioperative environment. Further, there are many and varied military applications from video summarization. The UMBC Graduate

students currently working on the CAST and the background research effort for the new direction will transition out, as the new direction is less closely aligned with their research interests. Dr. Mike Grasso, MD/PhD in Computer Science, will be joining the effort, and a new graduate student whose research focus will be on the video efforts will join the team.

Informatics subgroup 4. Operating Room Clutter (ORC)

The project team has worked on the use of advanced video technology to support coordination in operating rooms. Our activities were in four areas. All publications referred to may be found in our website: http://hfrp.umaryland.edu. For full length journal articles, PDF files may be downloaded. For others, abstracts are available. In all, we published 8 full-length peer reviewed journal articles, 2 full-length peer reviewed proceeding articles, and 8 conference abstracts. The references below can provide further details.

A. Models of decision making for operating room management.
We reviewed literature and developed a synthesis report on the state of the art of decisions on the day of surgery. Furthermore, we developed models for decision support systems for operating room management. The activities in this area were reported in the following publication:

1. Dexter F, Xiao Y, Dow AJ, Strader MM, Ho D, Wachtel RE. Coordination of Appointments for Anesthesia Care Outside of Operating Rooms Using an Enterprise Wide Scheduling System. *Anesthesia and Analgesia. 105:1701-1710. 2007*

B. Operating room multimedia system design and methodology.
We developed technology, primarily based on algorithms of video processing and biosignal processing, to display status of operating rooms. The displays are to increase situational awareness. The technological advances made by our group were reported in the following publications:

2. Xiao Y, Schimpff S, Mackenzie CF, Merrell R, Entin E, Voigt R, Jarrell B. Video Technology to Advance Safety in the Operating Room and Perioperative Environment. *Surgical Innovation. 14(1): 52-61. 2007*

3. Hu P, Xiao Y, Ho D, Mackenzie CF, Hu H, Voigt R, Martz D. Advanced Visualization Platform for Surgical Operating Room Coordination: Distributed Video Board System. *Surgical Innovation. 13(2):129-135. 2006*

4. Hu P, Seagull FJ, Mackenzie CF, Seebode S, Brooks T, XiaoY. Techniques for Ensuring Privacy in Real-Time and Retrospective Use of Video. *Telemedicine and e-Health, 12(2): 204, T1E1. 2006*

C. Survey and descriptive studies of operating room management, with and without the support of advanced video technology.

In conjunction with technology development, we conducted observational and survey studies of operating room management. These studies and associated results were in the following publications:

5. Seagull FJ, Xiao Y, & Plasters C. Information Accuracy and Sampling Effort: A Field Study of Surgical Scheduling Coordination**.** *IEEE Transactions on Systems, Man, and Cybernetics, Part A:Systems and Humans. 24(6), 764-771. 2004*

6. Dutton R, Hu PF, Mackenzie CF, Seebode S, Xiao Y. A Continuous Video Buffering System for Recording Unscheduled Medical Procedures. *Anesthesiology, 103:A1241. 2005*

7. Gilbert TB, Hu PF, Martz DG, Jacobs J, Xiao Y**.** Utilization of Status Monitoring Video for OR Management. *Anesthesiology, 103:A1263. 2005*

8. Dutton R, Hu P, Seagull FJ, Scalea T, Xiao Y, . Video for Operating Room Coordination: Will the Staff Accept It?. *Anesthesiology: 101: A1389. 2004*

D. Technology evaluation.

We conducted evaluation studies of the technology deployed. The primary focus was on user acceptance and usage patterns. The focus was chosen because the current science of operating room management has concluded that improvement of decision making on the day of surgery will lead to improvement in intangible outcomes, such as situation awareness, and will unlikely lead to improvement in operating room throughput (e.g., volumes and economic returns). Our work was reported in the following publication.

9. Xiao Y, Dexter F, Hu FP, Dutton R. Usage of Distributed Displays of Operating Room Video when Real-Time Occupancy Status was Available . Anesthesia and Analgesia 2008; 106(2):554-560. 2008

10. Kim Y-J, Xiao Y, Hu P, Dutton RP. Staff Acceptance of Video Monitoring for Coordination: A Video System to Support Perioperative Situation Awareness. *Journal of Clinical Nursing (accepted). 2007*

The project team has worked on the use of advanced video technology to support coordination in operating rooms. We developed models for decision support systems for operating room management. We developed technology, primarily based on algorithms of video processing and biosignal processing, to display status of operating rooms.

In conjunction with technology development, we conducted observational and survey studies of operating room management. We conducted evaluation studies of the technology deployed. The primary focus was on user acceptance and usage patterns. The

focus was chosen because the current science of operating room management has concluded that improvement of decision making on the day of surgery will lead to improvement in intangible outcomes, such as situation awareness, and will unlikely lead to improvement in operating room throughput (e.g., volumes and economic returns).

The Operating Room Clutter project enters its final phase under the provisions of the contract, and will end in February, 2009. Further research activities will seek support from other funding agencies.

Informatics subgroup 5. Improving Perioperative Communications (IPC)

Background:

In the UMMS OR the Cardiac Surgery Service utilizes a common communications point (a "cardiac phone line") that in a sense is used to acquire information and provide that information to any team member who calls the line to acquire information. The cardiac phone line has been scripted and is actively in use through a voice mail system. It can only be altered by dedicated personnel with password capability. The script involves the following standardized information: Identification of individual providing information, the Date of surgery, the Total number of cases, and OR location, patient name, case order, medical record number, age, surgeon, anesthesiologist and procedure. Evening schedule updates have been made possible through a second phone line option.

Problem Statement:

After some effort, we can now move to track updates on the phone line and correlate these updates with OR start delays. Thus, we refined the IPC question to Does more accurate information as evidenced by updates on the phone line, ie improved communication, result in fewer problems in the morning with cardiac surgical cases starting on time- are instruments better prepared for the procedure, are operating rooms better equipped for the appropriate case, are the correct pick lists utilized for the correct surgeon, is there less of a transport delay because the patient's hospital location has been identified? The question contains reference to some of the delay codes that are currently utilized by the Operating room tracking system and reported for glitch analysis.

With the assistance of the communications personnel reconfigured the cardiac phone line so that we can actually track the phone calls made to the phone line. This enabled us to: Determine key personnel who are utilizing the phone line, Determine groups of personnel utilizing the phone line (i.e. nursing, anesthesia, perfusion), Determine which groups are not utilizing phone line information (i.e. anesthesia techs), Determine whether there is a time variable; is there a better time to call for updates? Should updates be made at predetermined times or should they be more dynamic?

We hypothesized that information gained from increased communication improves OR efficiency. If this is the case we can them move to see if more real-time enabling

technologies might be deployed to other services within the UM ORs and perhaps other ORs “everywhere”.

During this period of performance, the team sought to find an appropriate “question” with which to focus this effort. In particular, there was a need to tie a performance metric (perioperative workflow related) to the IPC task. Although some progress was made, this aspect of Innovations in the Surgical Environment will end and future efforts will be refocused. Funding will be sought from other agencies

B. Simulation

Simulation.1 The Maryland Virtual Patient

We present here a simplified description of the MVP simulation, interaction and tutoring system. A virtual patient instance is launched and starts its simulated life, with one or more diseases progressing. When the virtual patient develops a certain level of symptoms, it presents to the attending physician, the system’s user. The user can carry out, in an order of his or her choice, a range of actions: interview the patient, order diagnostic tests, order treatments, and schedule the patient for follow-up visits. The patient can also automatically initiate follow-up visits if its symptoms reach a certain level before a scheduled follow-up. This patient-physician interaction can continue as long as the patient “lives.”

As of the time of writing, the implemented MVP system includes a realization of all of the above functionalities, though a number of means of realization are temporary placeholders for more sophisticated solutions, currently under development. The most obvious of the temporary solutions is the use of menu-based patient-user interaction instead of natural language interaction. While this compromise is somewhat unnatural for our group, which has spent the past 20 years working on knowledge-based NLP, it has proved useful in permitting us to focus attention on the non-trivial core modeling and simulation issues that form the backbone of the MVP system.

MVP currently covers six esophageal diseases pertinent to clinical medicine: achalasia, gastroesophageal reflux disease (GERD), laryngopharyngeal extraesophageal reflux disease (LERD), LERD-GERD (a combination of LERD and GERD), scleroderma esophagus and Zenker’s diverticulum.

At the beginning of a simulation session, the system presents the user with a virtual patient about whose diagnosis he initially has no knowledge. The user then attempts to manage the patient by conducting office interviews, ordering diagnostic tests and prescribing treatments.

Answers to user questions and results of tests are stored in the user’s copy of the patient profile, represented as a patient chart. At the beginning of the session, the chart is empty and the user’s cognitive model of the patient is generic – it is just a model of the generalized human. The process of diagnosis results in a gradual modification of the

user's copy of the patient's profile so that in the case of successful diagnosis, it closely resembles the actual physiological model of the patient, at least, with respect to the properties relevant to the patient's complaint. A good analog to this process of gradual uncovering of the user profile is the game of Battleship, where the players gradually determine the positions of their opponent's ships on a grid.

At any point during the management of the patient, the user may prescribe treatments. In other words, the system allows the user not only to issue queries but also to intervene in the simulation, changing property values within the patient. Any single change can induce other changes – that is, the operation of an agent can at any time activate the operation of another agent.

Simulation: Utility

The MVP project can be viewed as just one of a number of applications in the area of intelligent clinical systems. The latter, in turn, can be viewed as one of the possible domains in which one can apply modeling teams of intelligent agents featuring a combination of physical system simulation and cognitive processing.

So, in the most general terms, our work can be viewed as devoted to creating working models of societies of artificial intelligent agents that share a simulated "world" of an application domain with humans in order to jointly perform cognitive tasks that have until now been performed exclusively by humans. Sample applications of such models include:

- a team of medical professionals diagnosing and treating a patient (with humans playing the role of either a physician or a patient)
- a team of intelligence or business analysts collecting information, reasoning about it and generating analyses or recommendations (with humans playing the role of team leader)
- a team of engineers designing or operating a physical plant (with humans playing the role of team leader)
- a learning environment (where humans play the role of students).

As can be seen, this work is at the confluence of several lines of research – cognitive modeling, ontological engineering, reasoning systems, multi-agent systems, simulation and natural language processing.

During the period of performance, we have been working on the following issues:

1. We have continued to develop a computational model of the cognitive agent. We have tested the goal- and plan-based reasoning component and its interaction with the interoceptive and language perception modules and verbal, mental and physical action simulation modules.

2. We have spent much of the time in May and June preparing for the demonstration of the system at the program conference on June 27, 2008. In particular, we have developed a new demo interface.

3. We have continued to work on the natural language substrate of the system, concentrating on enhancements required for processing dialog (not expository text). As part of this module, we have implemented an enhanced microtheory of indirect speech acts.

4. We have continued working reference resolution algorithms (this is a very difficult task in and of itself).

5. We have continued work on the acquisition of ontology and lexicon knowledge.

6. Improvement of the DEKADE user interface has continued apace. New facilities for editing and viewing intermediate and final results of text analysis have been introduced and existing ones improved.

7. We have spent considerable time on improving the documentation of the project work. We have written and submitted for publication 5 papers describing aspects of our system.

Simulation.2 Training for Surgical Excellence and Patient Safety

The development of the Maryland Advanced Simulation, Training, Research and Innovation center (MASTRI) opened the door to innovative research opportunities that enhance surgical training and improve patient safety. Within the existing scope of the current contract, several projects will be undertaken during the final year of the contract that conceive, develop and validate simulation-based training for proficiency in the performance of surgical tasks.

We will create an environment that replicates the operating room and critical care unit of a medical center for the purpose of conducting scenario-based training of responses to emergent and post-operative crises. We will structure a training site to strengthen the response skills of students, residents and attending staff by building scenarios for learning, training staff to operate equipment and communication systems and developing assessment tools and metrics to measure performance. This training site will serve also as the team training site for medical center and school personnel.

We will accelerate our program of model development for simulation training. We have demonstrated success in the creation of physical models for surgical training, including a model for hernia repair that has attracted the attention of commercial product developers. This experience coupled with the clinical expertise available within the medical center portends success in model development, with particular reference to orthopaedics, OB GYN, critical care and general surgery. The recent purchase of a 3D printer expanded our capability to craft tissues for simulation training.

We plan to develop performance metrics for the comparison of physical and virtual reality models for simulation training. A logical extension of the development of performance metrics is the establishment of collaborative research relationships with other academic medical institutions. Such is our intent.

C. Smart Image

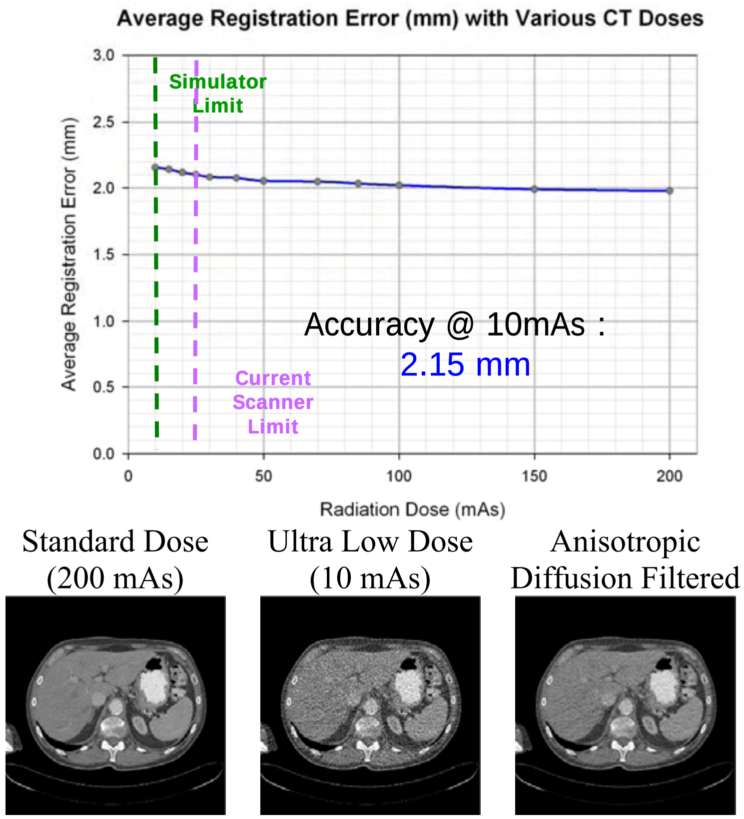
C.1. Smart Image: CT guided imaging

The overall objective of this project is to demonstrate the technical feasibility of live augmented reality (AR), which is the fusion of instantaneous computed tomography (CT)-generated views with laparoscopic views. The advantage of live AR is that internal structures that are absent in laparoscopic views can be visualized. Being able to see underlying vessels and other internal structures before making a dissection has been a longstanding need of the minimally invasive surgeons.

Although the proposed use of continuous CT is advantageous to creating renderings of the internal structures with their orientations updated at real-time rates, it is also imperative that methods be created to reduce the radiation exposure to the patient and the surgeon alike through the use of CT. Our first three objectives address the radiation dose problem while providing a means to visualize the vasculature throughout the surgery with a single administration of the CT contrast agent. The fourth objective is devoted to creating spatially and temporally synchronized AR views. The last objective will integrate all the individual technology solutions together when those are developed.

Objective 1: Dose reduction strategy: Registration between High dose-Low dose CT

The ultimate goal of this objective is to be able to track tissue motion through use of intra-operative low-dose CT scans and their deformable registration with peri-operative, diagnostic-quality CT scan. We have demonstrated the feasibility of intensity-based deformable registration between low-dose CT and regular-dose CT [1]. This preliminary study preformed registration between standard-dose CT (representing a pre-operative image) and ultra low-dose CT (representing intra-operative image). Even at 10 mAs, the smallest dose achievable, the registration accuracy achieved was comparable to that achieved at the standard dose. These results (see figure below) demonstrate the potential for ten- to twenty fold reduction in radiation dose with the use of low-dose CT. We are currently working on extending this study using human data.



Pre-processing of Low-dose CT

One of the challenges of working with low-dose CT is the poor quality (signal-to-noise ratio) of these images. We have developed image preprocessing techniques to eliminate this noise and the make these image more suitable for intensity-based deformable.
image registration. The results of these preprocessing techniques are shown in the following figure. For on-line use, however, these pre-processing steps must be accelerated such that the pre-processing operations can be performed in real-time. We have developed and reported [2] an FPGA-based implementation that provides over an order of magnitude speedup and is capable of performing this preprocessing in few milliseconds.

In addition to the developments mentioned above, we have also attempted to validate the feasibility of the above approach using animal datasets. This was achieved by using the low-dose and regular-dose CT data collected during the animal experiment. However, due to a localized deformation created using an ad-hoc method (syringe-pump), the deformable registration algorithm did not produce expected results. We investigated this issue and devised strategies to introduce a controlled and reproducible deformation. For further testing of these ideas we performed an experiment using a chicken, which allowed us to introduce reproducible deformations in a controlled fashion. These deformations were then track through the use of intra-operative low-dose CT. The low-dose CT data collected during this chicken experiment was processed. We registered this data with pre-operative regular dose scans and the results indicated visually acceptable quality of

registration. This demonstrates the preliminary feasibility of deformable registration with low-dose CT in animal datasets.

Objective 2: Dose reduction strategy: Iterative reconstruction

Iterative reconstruction techniques are known to give better quality image reconstructions in the presence of noise. Hence these techniques are better suited for reconstruction of low dose images. Over the past year we have implemented the maximum Likelihood Expectation Maximization technique for reconstruction of images.

The algorithm has given better quality images for simulated low dose as well as normal dose data sets. However, since the MLEM is an iterative technique with typical reconstruction involving about 40 iterations, and each iteration involving one forward and one back projection, the algorithm is not very attractive for real-time processing over a uniprocessor.

To enable faster implementation of this computationally intensive algorithm, a parallel version of the same has also been implemented which can be executed on a cluster of multiple processors. The cluster based approach gives almost linear speed-ups as compared to a uniprocessor implementation. Table 2.1 below indicates the speed-ups achieved for a representative data set.

Xeon,3.6GHz,1.5GB RAM	1 core of a Quad-core Xeon node, 3.2GHz, 4GB RAM	32 cores on 8 Quad-core Xeon nodes, 3.2 GHz, 4GB RAM
180 sec	143 sec	5.54 sec

*Table 2.1 Iteration time for a 256*256 CT slice with 367 detectors and 360 projections*

The reconstructed image quality was the same irrespective of the amount of parallelization. To further speed-up the rate of convergence, an Ordered Subset version of the Maximum likelihood algorithm was also implemented. This version was also parallelized and can now be run on a cluster of multiple processors. The speed-up achieved was linear to the number of subsets used in the algorithm.

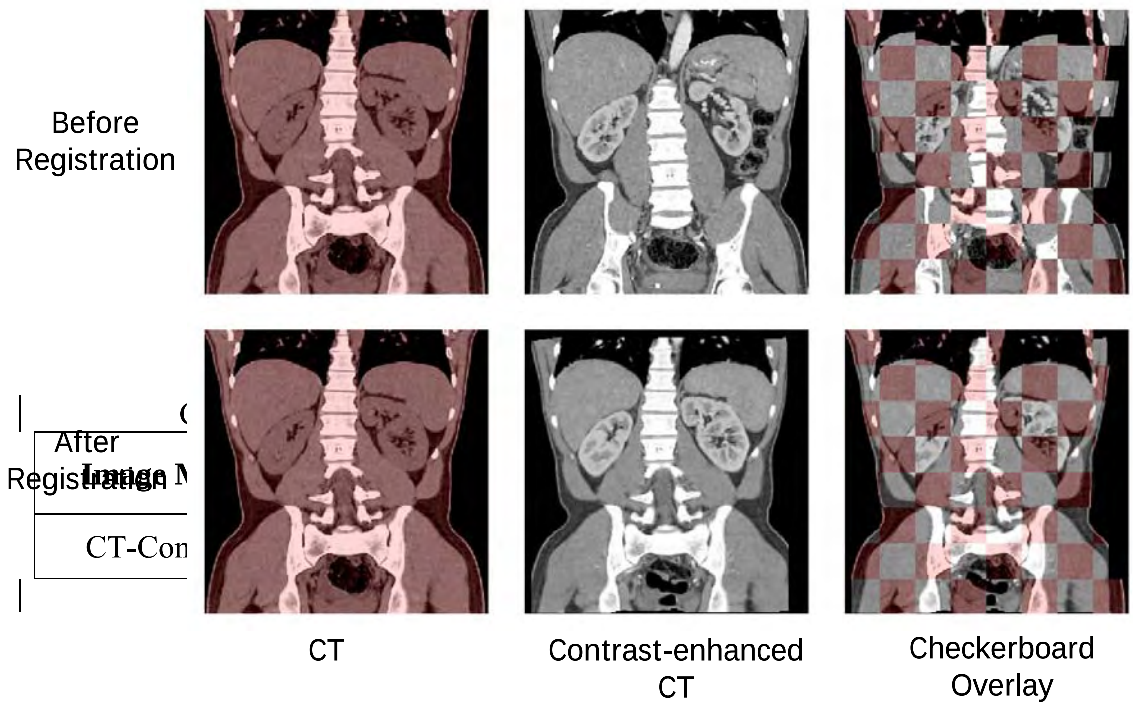
The implementation of the MLEM as well as OSEM algorithms is now being optimized to be more memory efficient to enable quick reconstruction of real data sets. To this end we are currently working on a low memory imprint version of the OSEM algorithm as well as a GPU based implementation of the same algorithm.

Objective 3: High-speed implementation of non-rigid registration

Deformable registration between intra-operative images and peri-operative images is a fundamental need in image-guided procedures. This registration will allow the fusion of complimentary information such as spatial information and vasculature information from pre- and peri-operative images respectively. Computational complexity of deformable

registration, however, has prevented its use in clinical applications. This objective attempts to address this problem through use of hardware-acceleration.

We presented the initial architectural design for accelerated deformable image registration [1]. This architecture is capable of calculating mutual information, a compute intensive step in intensity-based image registration, around 40-times faster than a software-based implementation can reduce the execution time of deformable registration from hours to minutes. The detailed design of this architecture was later published in IEEE - Transactions on Biomedical Circuits and Systems [2]. This architecture has been validated for CT-CT registration and will allow deformable registration in a matter of few minutes. The validation was performed using 5 CT-contrast-enhanced CT image pairs and the results of registration are presented in the following figure. Also, the accompanying table compares the execution times of this architecture against a software



implementation and reports the achieved speedup.

As a first step towards implementation of deformable image registration, we completed the hardware implementation of mutual information (MI)-based rigid registration. This implementation has been fully tested and rigorously validated using clinical and artificially deformed datasets. The accuracy achievable through this implementation is comparable to that achieved by a software implementation. This implementation is capable of providing 40-fold speedup in MI calculation and can achieve rigid registration is 50 seconds. Currently, we are working on optimizing this implementation for accuracy and hardware resources and a manuscript based on this work has been submitted to IEEE Conference on Field-Programmable Custom Computing Machines [3]. Once this optimization is complete, we will finalize and test the hardware implementation for deformable registration.

Objective 4: Tracking and visualization

We have developed the mechanism to track the laparoscope and other tools optically. Image 4.1 shows our experimental setup for tracking of operative tools.

Image 4.1: Experimental setup for optical tracking of intra-operative tools.

2 experiments were conducted over the duration of the last year to collect data and to improve the accuracy and the calibration methods used in the data tracking and collection methods. Each of the experiments has yielded progressively improved results. Some of the major improvements in the tracking and visualization methods due to the experiments are as follows:

1. We have moved to automatic volume rendering of the CT data for fast automatic 3-D volume generation as against the previous segmentation techniques.

2. Camera calibration has been integrated as a crucial step in the calibration procedure to account for the optical characteristics of the endoscope.

Image 4.2: Camera calibration plate with optical trackers

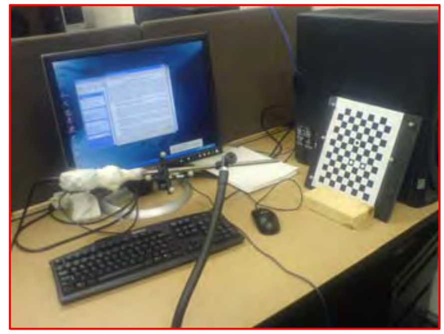


Image 4.3: Camera calibration setup

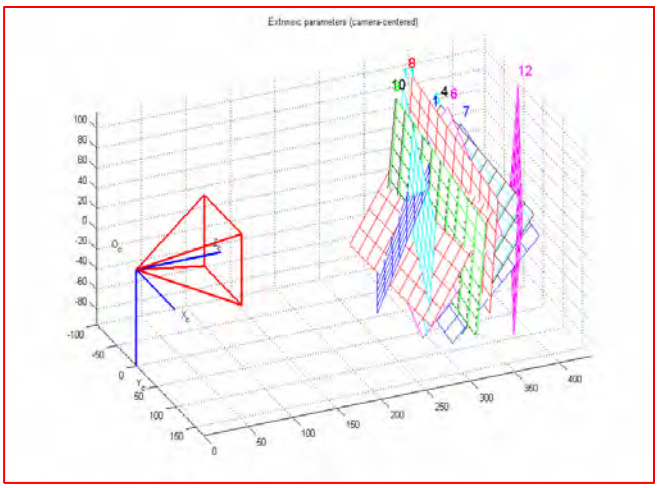


Image 4.4: Various positions of the calibration plate for camera calibration

3. Most of the steps in the tracking and visualization procedure have been completely automated to achieve fast and efficient registration of the CT reconstructed images with the optical images.

Though the experimental results have been improving, we still are facing some issues with the spatial and temporal registration of the optical images from the laparoscope and the 3-D volume rendered images from the CT scans. We are therefore working on isolating the unknown parameters in the tracking and visualization setup and devising improvements in our calibration techniques to overcome the problems.

Challenges

While our current efforts will prove the feasibility of live AR, implementing a real-time system for routine clinical use will require meeting a few additional technical challenges. First of all, the current generation CT scanners remain ill-suited for the task. Further development of the CT technology will be needed to improve the speed of both scanning and reconstruction to make CT real-time. We are developing a relationship with the manufacturer of our CT scanners (Philips) to address this challenge.

The use of multi-vendor devices is a challenge in creating spatially and temporally registered live AR display. Different devices exhibit different latencies, which are being estimated experimentally currently. Knowing those, or perhaps eliminating those, will improve the accuracy of live AR.

Finally, image reconstruction and registration are computationally intensive tasks. Although our present will show considerable acceleration of both, further acceleration will be needed to create a real-time system. We believe our current work will provide the necessary impetus to form necessary partnerships and take on the described challenges in a follow-on phase. During the latest portion of the period of performance, work was conducted on the objectives of the study as follows:

Objective 1. Dose reduction strategy: High dose-Low dose. (Algorithmic developments; phantom and animal model study; validation)

One of the challenges of working with low-dose CT is the poor quality (signal-to-noise ratio) of these images. We have developed image preprocessing techniques to eliminate this noise and have developed and reported an FPGA-based implementation technique to accelerate the registration of this data with high dose data, to ensure near real-time generation of good quality images from the low-dose data sets.

We performed an experiment over a live porcine model to collect low dose and contrast enhanced high dose data in June. The low-dose to high-dose registration for the data collected is being performed and we see acceptable initial registration results. We are using both surface based external markers as well as markers embedded in the liver to verify the registration results.

Objective 2. Dose reduction strategy: Iterative reconstruction (algorithmic developments; testing with scanner; software parallelization; hardware parallelization)

Iterative reconstruction techniques are known to give better quality image reconstructions in the presence of noise. We have previously reported accelerating the MLEM and PS algorithm on a cluster of computers. During this quarter, to ensure high speed reconstruction at low radiation doses at low hardware cost, we have used a general purpose GPU as a co-processor to accelerate the reconstruction process.

Using the low dose raw sinograms collected during the porcine experiment conducted in the last quarter, we have been able to quantitatively as well as qualitatively prove improvement in reconstructed image quality using the iterative PS algorithm. The GPU based implementation of the algorithm has helped to accelerate the algorithm by over 2 orders of magnitude.

Objective 3. High-speed non-rigid (hardware implementation; software parallelization; hardware parallelization)

We have previously completed the hardware implementation of mutual information (MI)-based rigid registration. This implementation has been fully tested and rigorously validated using clinical and artificially deformed datasets.

We continue to work on optimizing this implementation for accuracy and hardware resources. Once this optimization is complete, we will finalize and test the hardware implementation for deformable registration.

Objective 4. Tracking and visualization (Integration of tracking system; tool tracking for metal artifacts; automatic segregation; visualization; integration of visualization with tracking)

We conducted a pig experiment to verify our tracking and visualization techniques. However, we suffered a slight set-back due to hardware failure during the experiment. As a result we were unable to collect all the data that was essential for complete validation of our system. Since then, the hardware has been repaired and tested. We plan to conduct another experiment in the month of July to verify the efficacy of our techniques.

Objective 5. System integration:
System integration is one of the last milestones of the project and will follow other milestones after they are met.

The abstracts of our work are presented in the Appendix to this report.

C.2. Smart Image: Image Pipeline

Multimodal Registration Development and Evaluation (Intra-operative)

During the past year, we have developed and evaluated a technique to address the classic multimodal registration problem. The goal of our technique is to allow the integration of multiple laparoscopic views into a unified (3D, wide field of view) image. The technique has the following characteristics:

a. It does not rely on a tracked camera (traditional approach). This is important because it is difficult to get accurate camera parameters using available tracking techniques. As a result, our evaluation of the traditional method produces obvious shifts between the target (desired) and actual (obtained) images.

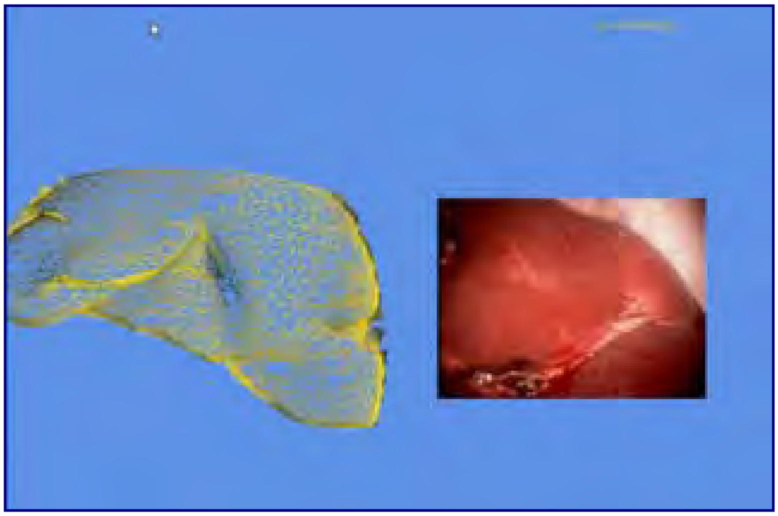
b. It can deal with deformable objects. This is important because of the deformations introduced to objects during surgical procedures.

c. It changes the difficult 2D-3D registration problem into a 2D-2D registration problem which is better understood.

Our approach begins from a single interactively aligned image/3D model pair. It involves an incremental global-optimization algorithm to automatically register

additional nearby images. We demonstrated a registered data set during the ORF 2007 conference. This example is presented below.

.



We have more formally evaluated the approach by testing the simplified case of a sphere surface. Compared with ground truth, the error was calculated by determining the average distance of texture coordinates per vertex (482 vertices, 960 faces, and 81 feature points, mean error per vertex < .0045).

We next turned to a test case for our registration technique that was considerably more complex than a sphere – a human face and a set of teeth. The results of this study were described in a paper, titled "Feature-based Texture Mapping from Video Sequences," that was accepted as a poster presentation at the Symposium on Interactive 3D graphics and Games (I3D 2008). The full paper is attached to this report.

Example of feature-based texture mapping applied to a face model to create a 3d image that retains surface detail.

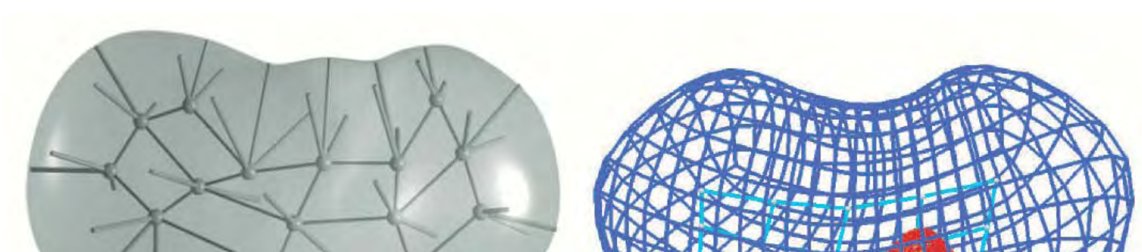
Our most detailed evaluation of our registration technique is currently in progress. We are using as stimuli anatomical models of abdominal organs. 3D data are being obtained by scanning the organs with a Faro Arm (platinum model), a 3-d scanner that uses the position of a fully-articulated arm with a laser-scanner mounted on the end to gather data used to create a 3D model. The surface images of the organs are being acquired using an attached Stryker 1088 endoscope. For comparison purposes, registered output can be generated using the relatively precise motion tracking of the Faro Arm in order to obtain location parameters of the attached endoscope. This will provide us with a baseline condition that utilizes the traditional camera tracking registration methodology with our new technique. In order to make sure that our tracking data are as accurate as possible, a special mount was built to enable a more rigid attachment of the endoscope to the arm.

Initial tests using the Faro arm and attached endoscope have shown promise. Using previous work developed under the REVEAL project, the endoscope's calibration parameters have been extracted from a video sequence of a simple calibration target. These parameters are then combined with the tracking data from the Faro arm to calculate the transformation from the camera to the arm. Software to merge this tracking data with the high-definition video sequence for structure-from-motion 3D reconstruction is also currently in production by Dr. Yang's team

Multimodal Registration Development and Evaluation (Pre- and Intra-operative Data Integration)

In addition to exploring methods for integrating data from multiple intra-operative sources, we have been investigating methods for registering pre- and intra-operative data. This effort has been led by Dr. Qiong Hon, a medical imaging specialist, who is the newest member of the Smart Image team. Our approach is based on the goal of computational efficiency, and to that end we are attempting to register shape models from different intra- and pre-operative modalities into the 2D view rather than attempting to register the images themselves.

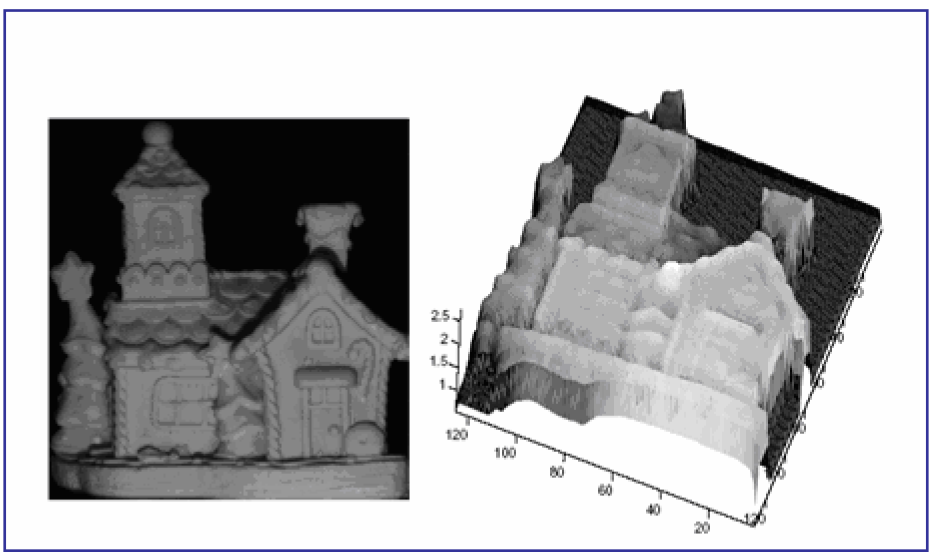
We are using a medially based shape representation called “m-rep” as our shape model. The advantage of an m-rep model lies in its full volumetric parameterization of both the object interior and the adjacent exterior and in its power to capture anatomical shape variations via its non-linear shape parameters. The m-rep has been shown to provide one of the best prostate segmentation results from CT images. Because the m-rep provides a full 3D volumetric shape model, certain manually or automatically extracted features such as landmarks or contours can be highlighted effortlessly. This provides the input for visualization techniques motivated by cognitive ergonomics principles, such as nonphotorealistic rendering, or “decluttering.” As described below, we have used the m-rep approach, along with intra-operative texture mapping, to produce a prototype “dual display” visualization that presents anatomical panoramas, detailed camera views, and the relationship between the two simultaneously.



Left: a deformable m-rep shape model; right: a tumor (red) in a kidney

A Novel Intra-operative Depth Acquisition Technique: Light Fall-Off Stereo

The Concept. We have developed a new method for depth acquisition during surgery that takes advantage of the enclosed, light-controlled environment of the surgical site. The technique, dubbed light fall-off stereo (LFS), uses the known properties of light attenuation as a function of distance to infer depth. Our initial results were presented at the International Conference for Computer Vision and Pattern Recognition (CVPR) – one of the premier conferences in the computer vision community. This paper is available at http://vis.uky.edu/~wangl/Research/Publication/2007/Light_fall-off_stereo.pdf. We built a real-time prototype that was presented at the ORF 2007 Conference, examples from which are presented below. While the initial results are quite encouraging, we still need to evaluate its performance with more experiments.



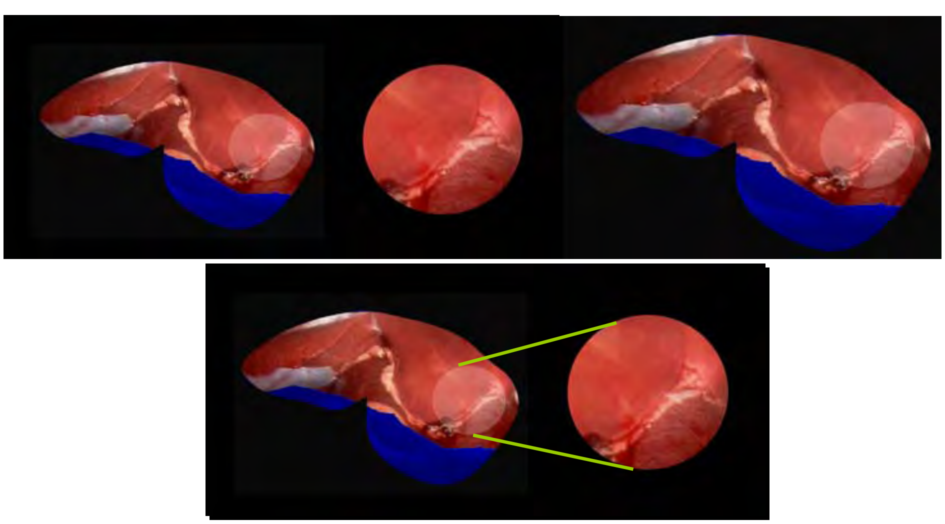
Potential Industry Partnership. Initial results of our work with LFS were successful enough for us to pursue the potential development of a prototype laparoscope that uses our new concept with Stryker Endoscopy. After discussing the project in an initial conference call on 9/27/07 with Stryker marketing representatives, we forwarded

relevant research papers and an executive briefing to their engineering department. Stryker showed sufficient interest in our approach to send three representatives, led by Brent Ladd, to visit the UK Center for Visualization and Virtual Environment for a demonstration on 12/5/07. All three Smart Image PIs (Carswell, Seales, and Yang), were present at the meeting. In addition to the demonstrations, we discussed the nature of the light source that would be required for the technique to be effectively incorporated into an endoscope. The Stryker team felt that the technique showed enough promise to invite Dr. Yang to Stryker headquarters for a discussion with its engineers. Stryker offered the first week of March as the time for a visit. Unfortunately, this resulted in a scheduling conflict for the UKY team. We are still negotiating a time to visit.

Smart Imagery Visualization Frameworks: Development and Evaluation

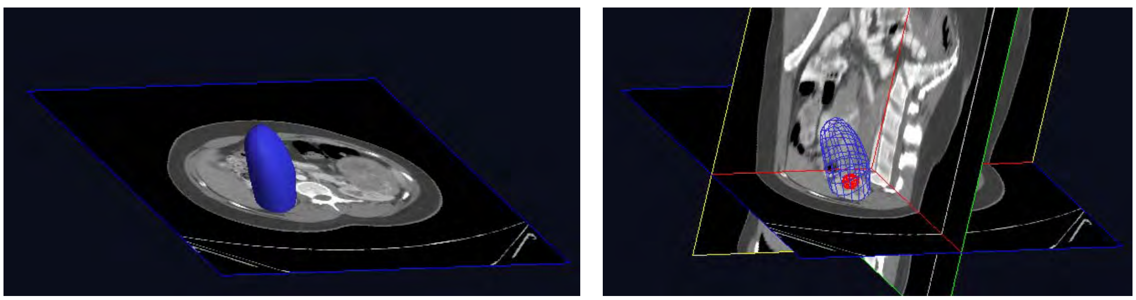
Display concepts. We have continued to pursue the development of visualization techniques that are motivated by cognitive ergonomics principles such as 1) exploitation of redundancy, context, and expectancy, 2) reduction of information access effort, and 3) reduction of memory loads. One such visualization technique is our “dual-view” display. We render the pre-built 3D (m-rep) shape models (described above) along with the original camera view in one of three ways, each method differing in the level of visual integration provided to the user. The shape models are texture-mapped using the panorama textures fused from video sequences of the camera view. In order to link the panorama rendering of the 3D shape models to the original camera, we use the tracked camera pose information to highlight the corresponding portion in the panorama view. In the “separate” dual-view display, the panorama and the camera view are provided in separate windows, with the approximate location of the camera view shown as a circular, highlighted area against the panorama. In the “connected” dual-view display, the panorama and camera views are still in separate windows, but now they are visually tethered by added contours. In the “integrated” dual-view display, the camera view is superimposed in its approximate location on the panorama itself.

All three dual-view displays show both the concentrated camera view with very limited field of view and the panorama with wide field of view and relate the two views to one another. We have described in more detail the motivation and development of this display in a proposal submitted to CARS 2008 (Principled Display Environment for Computer Augmented Minimally Invasive Surgery, Q. Han, M. Carswell, W. Wang, R. Yang, and B. Seales, Computer-Aided Radiology and Surgery (CARS) 2008.).



Three modes of the dual display environment. Top left: separate mode, top right:: integrated mode, and bottom: linked mode.

In addition to the dual-view displays, we have developed mock-ups of various "hierarchical" displays that highlight different part-whole structures. The "parts" include cross-sections highlighted within the global display, landmarks (e.g., embedded tumors, vascular structures, etc.), strata, metabolic activity gradients, and the like. Each part can be highlighted or removed according to the needs of the surgeon.



Demonstration of the rendering of different parts (information) in our hierarchical displays. Left: a kidney model (in blue) with a tumor sub-model embedded in one axial image slice; right: the same model with the tumor (in red) shown against the axial, sagittal and coronal image slices of the pre-op CT image from which the models are built.

Evaluation. We have developed an evaluation plan that leverages our recent work on situation displays designed for incident commanders of emergency response teams. These commanders, like surgeons, must work with inherently spatial data in safety-critical situations. The assumption behind this work is that different display formats, such as those described above, have differing degrees of "cognitive fit" or compatibility with different tasks. Ideally, the optimal match between tasks and well-designed display alternatives will be self-apparent. That is, the user should be able select the most appropriate display to use for different situations with little additional effort (i.e., without increasing mental workload). To determine whether this is possible, we must design different tasks that are best performed with different displays. That is, we must have normative task-display pairings. These could include, for example, judgments about tumor volume vs. judgments about tumor diameter at the tumor's widest point. A global view should be best for the first task, although a series of cross sections should be better for the second. There should be a large array of "correct" task-display pairings that best support the responses of users. An effectively designed set of rendering options (and these may be conceptualized as the result of user interactions with the "full" global model) will allow rapid and accurate judgments of appropriateness of the various formats

by users. So, does Participant X “pull” a cross-section from the global model to solve a problem for which it is the best-suited format?

We have begun development of the experimental scenarios at the same time that Dr. Cindy Lio, our lead usability analyst, is performing a more traditional needs assessment using our preliminary mockups as focal points of semi-structured interviews with surgeons. We will transcribe the interviews and analyze the transcripts to find emergent themes to gain insights about the surgeons’ perspectives on the visual displays.

In addition to the mock-ups for our displays described thus far, we have begun development of a more dynamic system for producing prototypes for higher fidelity user studies. Specifically, Dr. Han is developing a simulation program based on the hierarchical display environment. The FARO robot arm is used as an accurate and reliable tracker to control a virtual camera. The virtual camera pose is then used by the simulation program to generate a simulated laparoscopy camera video sequence. In the simulation both synthetic objects and objects from the real world can be used as the targets, and the ground truth of both the object geometry and deformation field is available.

In addition to the performance tests described above, we will be able to utilize a new eye-tracking system to document display utilization at a more detailed level. We are currently validating the eye tracking system with traditional laparoscopic displays, documenting changes in display utilization as training progresses.

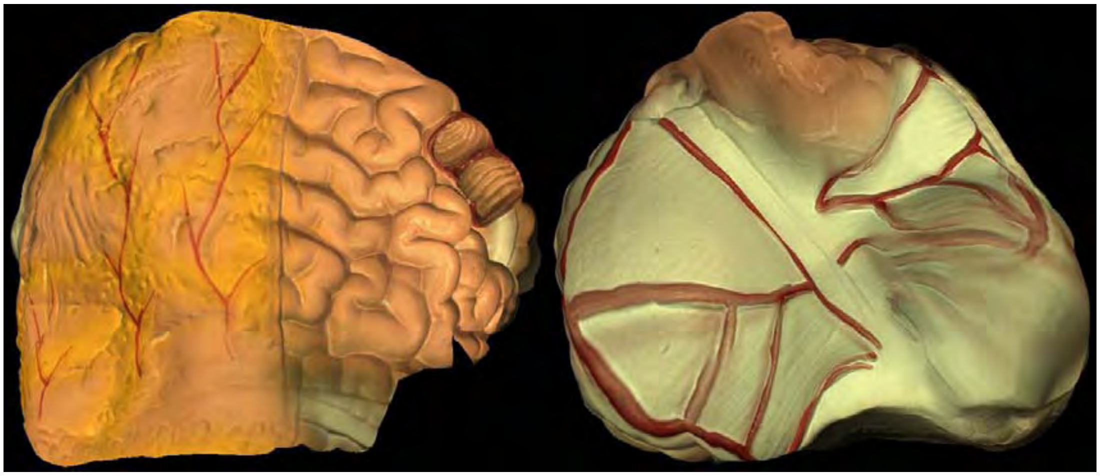
The faceLAB eye-tracking system tracks eye and head positions in real-time with a calibrated pair of stereo cameras. Tracking information from the eyes can be overlaid on video streams to accurately show what objects attract a subject's attention. Other useful metrics for cognitive ergonomics include blink rate, eye closure, saccadic movement, and pupil dilation. Together, this information can provide feedback on a subject's stress, mental workload, and fatigue, giving both the psychologists and the software developer an objective measure of the effectiveness of their work to the end user.

Summary of Progress

During this period of performance, we have worked primarily on 1) testing our texture mapping technique for adding scope-view data to predetermined 3-d models, 2) getting surgeon feedback on the current simulated “dual display” framework, and 3) developing further the dual display framework in order to perform user testing. Although these have been our primary goals, we have also been continuing our efforts to refine our cognitive ergonomics measures.

Texture Mapping Technique: Evaluation

Over the past year, we have been evaluating our method of registering scope-view data to predetermined 3-d models, a technique that is innovative in that it does not rely on knowledge of camera parameters, information that is sometimes difficult to obtain at levels of acceptable accuracy. During the first quarter, we collected video segments from anatomical training models to demonstrate our technique and to use in our formal evaluation of registration accuracy. However, we were faced with technical challenges when developing the baseline against which to compare our new technique. Specifically, the Faro Arm used for tracking camera pose was either broken or off site for extended periods, and we also had to develop a method of rigidly mounting our Stryker 1088 endoscope to the tracking arm. We have resolved those problems in the past quarter, and have conducted our evaluations, finding that our method results in less registration error than the traditional method. These data have been written up in a paper submitted to IEEE Trans.: Medical Imaging. The following are two views of a texture mapped intestine.



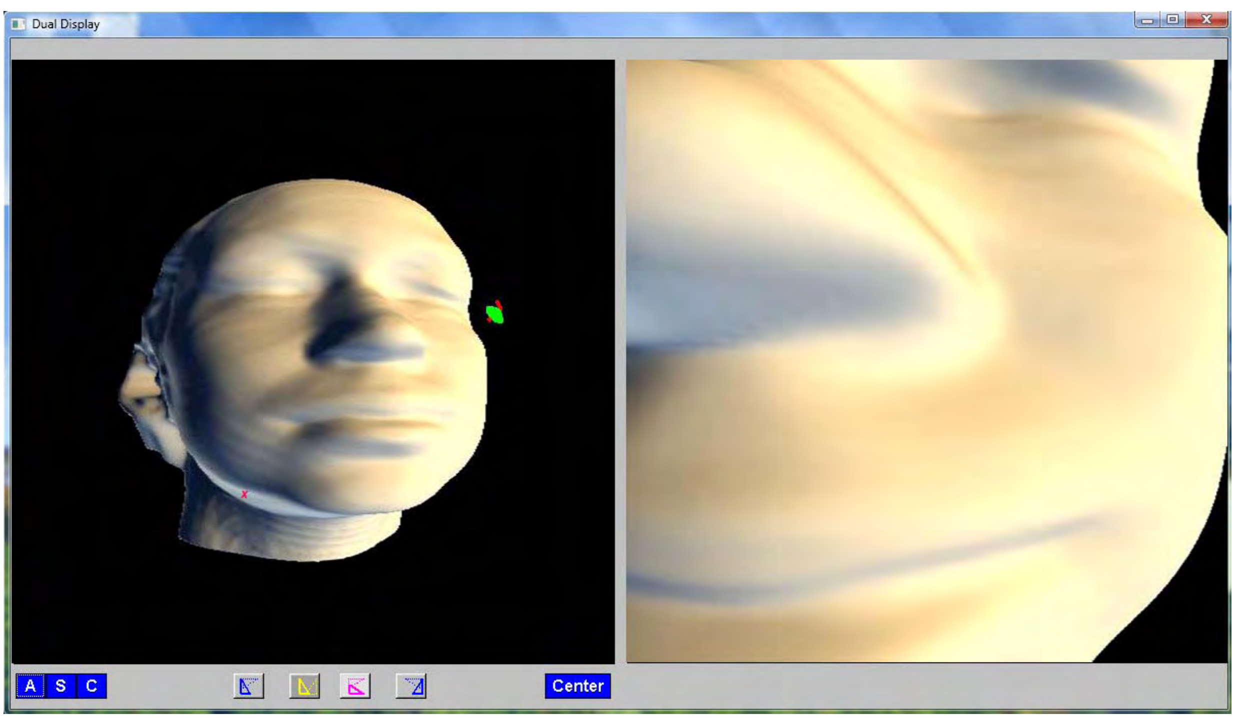
Extending Methods to Deformable Objects

Our current method assumes a rigid 3D model is known a priori (e.g., from pre-op imaging). Our plan is to extend this method to handle deformable objects without requiring the knowledge of a 3D model as input. The basic idea is to use machine learning methods to find a mapping between the 2D appearance and its 3D structure of a 3D model. We will first collect a training data set of a deformable model with known 2D to 3D correspondences. This can be obtained via simulation data. Then non-linear dimension reduction techniques can be applied to automatically find the mapping. After

the training is done, we will evaluate its effectiveness with real data. Ultimately the goal is to create complete 3D models from a monocular video sequence.

Dual Display Framework

At the same time that we are developing better ways to create 3d models, we are also trying to determine the most effective ways to view the information, with special emphasis on integrating local (scope view) and global (3d panoramas) to optimize performance in a variety of surgical navigation tasks (e.g., navigation of the scope itself, navigation of instruments, searching for target "terrain"). To this end, we have proposed a dual display environment, in which the scope view and the global view are rendered side by side or, in some cases, with one superimposed on the other. In order to test the effectiveness of various configurations we have begun by getting feedback on an early prototype. Thus, surgeons at the 2008 SAGES conference were invited to provide feedback. In addition, we will begin testing research participants using a human face as the target object. We will initially obtain quantitative evaluations from nonsurgeons performing a simple search task, where an 'x' will be the target and the participants will be using as a baseline only a local scope view to scan the face. A face was chosen as the search stimulus because familiarity with structures can greatly increase search efficiency, and few of our initial subjects will be suitably familiar with internal organ structures. The relative location of our component displays, color coding, and method of integration can all be tested to determine differences in search time, search efficiency, and perception of mental workload. A show of one dual display format is shown below for the face data.



The left panel shows the global view, with the icon above the cheekbone indicating the simulated location of the scope. The right panel shows the "scope view" that would result from this camera orientation. The participants will experience control

of the camera location by manipulation of the 3-d stylus of a Phantom desktop haptic interface.

D. Ergonomics and Human Factors.

Recently, a fourth pillar was added to our research portfolio, that of Ergonomics and Human Factors. These are two related branches of study that examine the relationship between people and their work environment. Ergonomics often focuses on the physical environment and the human body, while human factors center more on the cognitive aspects of performance. The same ergonomics and human factors techniques credited with making industrial processes safer and more efficient can be applied to the analysis and improvement of OR operations. Tools, such as video analysis and motion tracking, can be used to analyze current practices, identify inefficiencies and dangers, develop solutions, and measure improvement. "Best practices" to maximize safety and efficiency can be developed based on empirical data.

Our discussion of workflow to this point has taken a macro or panoramic view; for example, how might we most effectively track and bring together the people and assets necessary to ensure that a patient's surgical experience is safe and efficient. Through human factors and ergonomics, we have the ability to focus on a more micro-level analysis, such as how the physical interface between the surgeon and the patient could be improved and the associated work space chaos and stressors of MIS be reduced.

The patient is the center of the ORF. During MIS, the interfaces between the patient and the surgeon are critical to both the safety and quality of patient care and surgeon welfare. Patient-surgeon interfaces are complicated by compromises in equipment design, technology limitations, operating theatre layout, and technical approaches. In particular, ergonomic problems in the MIS workspace, such as obstructing catheters and cluttering tubes, can elevate the chance for contamination, increase surgical risks to the patient, and reduce work efficiency. Optimal workflow during MIS stands to be achieved through better understanding of patient-surgeon interfaces, both intracorporeal and extracorporeal. In the ORF, advanced technology could function as a key enabler, allowing an optimal patient-surgeon interface.

Some of our current work is focused on establishing quantitative, valid measures of workflow within patient-surgeon interfaces, identifying ergonomic problems that result as a consequence of workplace designs (e.g., arrangement or management of cables and catheters), and demonstrating key barriers to optimal workflow that present direct safety and efficiency concerns. One project is based on collaboration between surgical experts and human factors experts. Previous experiences in video capturing and analysis are being used as a basis for development of workflow measures and identification of ergonomic inadequacies. Time-motion studies have been conducted to collect objective data on activities in the patient-surgeon interface. Conceptual workplace layout designs are being developed based on objective data and simulations of what workflow might be if interfaces were optimized.

Given the physical risks associated with performing laparoscopic surgery, ergonomics to date has focused on the primary minimally invasive surgeon. Similar studies have not extended to other operating room staff. Simulation of the assistant's role as camera holder and retractor during a Nissen fundoplication allowed investigation of the ergonomic risks involved in these tasks.

Several manuscripts related to our research in ergonomics and human factors are placed in the Appendix to this report.

Key Research Accomplishments

A. Informatics

Informatics subgroup 1. Perioperative Scheduling Study
Major Accomplishments achieved during this period of performance include the successful analysis of patient data for all surgical cases, the development of an algorithm for predicting bed requirements, and the comparison of a variety of prediction methods for two units.

Informatics subgroup 2. Operating Room Glitch Analysis
Data architecture related to operating room performance and delay has been integrated into the hospital information system, giving visibility of OR events to clinical and operational leaders.

Informatics subgroup 3. Context Aware Surgical Training (CAST)
A prototype CAST system was emplaced in the MASTRI system for assessment. Work was done to design a system of evaluation of the system in terms of improvements in learning outcomes due to self-feedback, improvements in learning outcomes due to instructor feedback and synchronous versus asynchronous feedback.

Informatics subgroup 4. Operating Room Clutter (ORC)
During this period of performance, we published 8 full-length peer reviewed journal articles, 2 full-length peer reviewed proceeding articles, and 8 conference abstracts.

B. Simulation (Virtual Patient)

In this year of the project, our team has delivered additional versions of the Maryland Virtual Patient Environment. The realism of the simulation has been enhanced by including coverage of "unexpected" interventions; allowing discontinued treatments; allowing new diseases to develop due to side effects of treatments. The user interface has been redesigned. A new agent-based architecture has been developed to support enhanced cognitive capabilities of the virtual patient and the intelligent tutor, including language capabilities. In the area of language processing, a dialog processing model was developed. Work has continued on improving the language understanding capabilities, centrally including treatment of referring expressions. Enhancement of static knowledge

resources, the ontology and the lexicon, has been ongoing. Work on extending the coverage of diseases has been ongoing: a further improvement of the model of GERD is under way, as is the modeling of cardiovascular diseases. A totally reworked system version, with dialog support, is planned for release in June 2008. Work has also been ongoing on improving and extending the set of development tools – the DEKADE demonstration, evaluation and knowledge acquisition environment supporting natural language work has been revamped; the interface for creating instances of virtual patients has also been enhanced; a web-based environment for supporting internal documentation has been installed.

C. Smart Image

C.1. Smart Image: CT guided imaging

For Objective 1: Dose reduction strategy: Registration between High dose-Low dose CT, we demonstrated the feasibility of deformable image registration with low-dose CT, and demonstrated the potential for up to 20-fold reduction in radiation dose. For Objective 2: Dose reduction strategy: Iterative reconstruction, we developed iterative reconstruction algorithm for reconstruction of low-dose CT images. This implementation offers improved image quality at low-dose when compared with scanner-based reconstruction. Also, we implemented a parallel version of this algorithm that offers about 25-fold speedup and provides same image quality.

For Objective 3: High-speed implementation of non-rigid registration, we designed and developed an FPGA-based architecture for accelerated implementation of deformable registration algorithm. This architecture is capable of providing 40-fold speedup for image registration. We achieved also implementation and validation of intensity-based rigid registration using the aforementioned architecture and demonstrated a capability of performing rigid registration (first step to deformable image registration) under 1 minute.

For Objective 4: Tracking and visualization, we developed the mechanism to track the laparoscope and other tools using optical tracking and developed the core components of a visualization system to provide live-augmented reality.

.

C.2. Smart Image: Image Pipeline

Our accomplishments during this period of performance included these. We developed a new method of intra-operative registration that relies on feature-based texture mapping to spatially integrate video sequences into panoramas. We performed and published initial technical evaluations on the new registration technique, and developed a method of producing “baseline” registration samples using more traditional (camera tracking) procedures against which to compare our new technique.

We developed a new method of intra-operative depth acquisition dubbed “light fall-off stereo”, initiated collaboration with Stryker Endoscopy to integrate light fall-off stereo into a prototype endoscope, developed design concepts for visualization techniques based on principles of cognitive ergonomics. We began evaluation of the design concepts using

semi-structured interviews with surgeons and began development of stimulus sequences for use in performance-based user testing.

Reportable Outcomes

We advanced the body of knowledge pertaining to informatics, smart image, simulation and human factors as these relate to surgical procedures, the perioperative environment and the training of surgery. We published XXXX manuscripts, hosted national and international meetings related to innovation in the surgical environment, and incorporated technical advances into patient care in a large academic medical center. We influenced significantly the training of more than three hundred fellows and residents, hundreds of staff and care providers and numerous medical students.

Perhaps our most important accomplishment has been the identification of a new of basic surgical sciences. These include computer and physical sciences, informatics, smart imaging, simulation and ergonomics and human factors that underpin surgical training. This event is a landmark of sorts, as it has changed forever the course of surgical education. Lessons learned from this research effort are being applied in training programs throughout the country and internationally.

Conclusion

This report began with the recognition that an extraordinary evolution in surgical care has occurred caused by rapid advances in technology and creative approaches to medicine. The increased speed and power of computer applications, the rise of visualization technologies related to imaging and image guidance, improvement in simulation-based technologies (tissue properties, tool-tissue interaction, graphics, haptics, etc) have interacted to advance the practice of surgery. However, the medical profession lags behind other applications of information systems. The research program reported here has proceeded under the mantle of "Operating Room of the Future". As a natural occurrence in the outcome of lessons learned in medicine, we are replacing that theme with the more appropriate "Innovations in the Surgical Environment."

This research program has consisted of three major pillars; OR informatics, simulation, and smart image. This year, we added a fourth pillar, that of physical and cognitive ergonomics and human factors as these impact the surgical environment.
\
The purpose of the OR informatics program is to develop, test, and deploy technologies to collect real-time data about key tasks and process elements in clinical operating rooms. We have established testbeds of activities in both simulated and operational environments. We are currently performing tests of the hardware, refining software, and applying lessons learned to hospital operational functions. The objective of Simulation research is to create a system where a user can interact with a virtual human model in cognitive simulation and have the virtual human respond appropriately to user queries and interventions in clinical situations, with a focus on cognitive decision making and judgment. We have made significant strides toward realizing these goals. The MVP

simulation functions well for esophageal disorders, and is continuing to expand the repertoire of diseases that are in the simulation model.

The objective of smart image is use real-time 3D ultrasonography and 40-slice highframe-rate computed tomography (CT) for intraoperative imaging to volume rendered anatomy from the perspective of the endoscope. We are combining CT and Ultrasound to overlay image and data to enhance the performance of surgeons-in-training. We have carried out animate model testing of the image registration with great success. We continue to refine and expand our capability through hardware and software refinement.

In the future, OR workspace layout would be optimized through ergonomic data and human factors analysis, and this optimization would lead to the establishment of "best practices" for an array of surgical operations. Proper layout would reduce risks of infection, speed operations, and reduce fatigue of surgeons and staff, all elements that could contribute to a reduction in AEs and improved patient safety.

The year ahead is full of promise for refinements in the use of informatics to support safe and efficient operating room procedures, the use of simulation to improve and accelerate the training of competent surgeons, and the blending of imaging capabilities to provide clearer and safer interactions between patient and surgeon.

As stated earlier in this report, the current contract, W81XWH-06-2-0057, has been tied to a prior and topically related contract, DAMD-17-03-2-0001. The prior contract will close in February of 2009; the current one in October 0f 2009. Some projects contained in the Informatics pillar, OGA, ORC and IPC, will terminate in February. The WORQ project will continue under the current contract as will the CAST project that has been reshaped into Video Summarization.

Under the Simulation pillar, the MVP will end in February; the surgical simulation project will continue until October. The Smart Image projects will continue until October as will the projects under the pillar of Ergonomics and Human Factors.

These changes represent the maturing of a research endeavor over the course of six years, an endeavor which opened the door to a new set of basic surgical sciences. The Innovations in the Surgical Environment conference planned for the summer of 2009 will summarize the entirety of the research effort and point the direction to future innovative approaches to advance surgical technology in behalf of patient safety.

Publications

ABSTRACT: High-Speech Reconstruction of Low-Dose CT Using Iterative Techniques for Image-Guided Interventions. Doctoral Dissertation, University of Maryland 2008. Venkatesh Bantwal Bhat

ABSTRACT: High-Speech Reconstruction of Low-Dose CT Using GP-GPU. Venkatesh Bhat and Raj Shekhar

ABSTRACT: Live Augmented Reality for Laparoscopic Surgery Using a Novel Imaging Method – Initial Results from a Porcine Animal Model. R Shekhar PhD, CD Godinez, M.D., S Kavic, M.D., E Hart, M.D., Ivan George, AE Park, M.D.

Nagy P, Ramon Konewko, Max Warnock, Wendy Bernstein, Jacob Seagull, Yan Xiao, Ivan George, and Adrian Park. "Novel, Web-Based, Information-Exploration Approach for Improving Operating Room Logistics and System Processes", Surg Innov 2008 15: 7-16.

Ordóñez P, P. Kodeswaran, V. Korolev, W. Li, O. Walavalkar, B. Elgamil, A. Joshi, T. Finin, Y. Yesha, I.George. "A Ubiquitous Context-Aware Environment for Surgical Training". The First International Workshop on Mobile and Ubiquitous Context Aware Systems and Applications (MUBICA 2007), August 2007.

Seagull FJ, Moses GR, Park AE. "Integration of Virtual Reality and Conventional Skills Trainers: A Mixed Resource Model." In Westwood J.D., Haluck R.S., Hoffman H.M., Mogel G.T., Phillips R., Robb R.A. and Vosburgh K.G. (Eds). Studies In Health Technology and Informatics Volume 132. Medicine Meets Virtual Reality 16 - parallel, combinatorial, convergent: NextMed by Design. pp. 446-50. 2008.

Moses GR and Park AE. "Ergonomic risk associated with assisting in minimally invasive surgery. " accepted in Westwood J.D., Haluck R.S., Hoffman H.M., Mogel G.T., Phillips R., Robb R.A. and Vosburgh K.G. (Eds). Studies In Health Technology and Informatics Volume 132. Medicine Meets Virtual Reality 16, 2008.

Dexter F, Xiao Y, Dow AJ, Strader MM, Ho D, Wachtel RE. "Coordination of Appointments for Anesthesia Care Outside of Operating Rooms Using an Enterprise Wide Scheduling System". *Anesthesia and Analgesia. 105:1701-1710. 2007*

Xiao Y, Schimpff S, Mackenzie CF, Merrell R, Entin E, Voigt R, Jarrell B. "Video Technology to Advance Safety in the Operating Room and Perioperative Environment". *Surgical Innovation. 14(1): 52-61. 2007*

Xiao Y, Dexter F, Hu FP, Dutton R. "Usage of Distributed Displays of Operating Room Video when Real-Time Occupancy Status was Available" . Anesthesia and Analgesia 2008; 106(2):554-560. 2008

Kim Y-J, Xiao Y, Hu P, Dutton RP. "Staff Acceptance of Video Monitoring for Coordination: A Video System to Support Perioperative Situation Awareness". *Journal of Clinical Nursing (accepted). 2007*

Dandekar O., K. Siddiqui, V. Walimbe, and R. Shekhar, "Image registration accuracy with low-dose CT: how low can we go?," in 3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006, pp. 502-505.

Dandekar O., C. Castro-Pareja, and R. Shekhar, "FPGA-based real-time 3D image preprocessing for image-guided medical interventions," Journal of Real-Time Image Processing, vol. 1(4), pp. 285-301, 2007.

Shetye A.S. and R. Shekhar, "A statistical approach to high quality CT reconstruction at low radiation doses for real-time guidance and navigation," Proc. SPIE Med. Imaging, 2007.

Dandekar O., V. Walimbe, and R. Shekhar, "Hardware Implementation of Hierarchical Volume Subdivision-based Elastic Registration" in 28th Annual International Conference of the IEEE: Engineering in Medicine and Biology Society, 2006, pp. 1425-1428.

Dandekar O. and R. Shekhar, "FPGA-accelerated Deformable Registration for Improved Target-delineation During CT-guided Interventions," IEEE Transactions on Biomedical Circuits and Systems, vol. 1(2), pp. 116-127, 2007.

Dandekar O., W. Plishker, S. Bhattacharyya, and R. Shekahr, "Multiobjective Optimization of FPGA-Based Medical Image Registration" IEEE Symposium on Field-Programmable Custom Computing Machines, Under Review, 2008.

Shekhar R., O. Dandekar, S. Kavic, I. George, R. Mezrich, and A. Park, "Development of continuous CT-guided minimally invasive surgery," Multimedia Meets Virtual Reality (MMVR), 2007.

Shekhar R., O. Dandekar, S. Kavic, I. George, R. Mezrich, and A. Park, "Development of continuous CT-guided minimally invasive surgery," Proc SPIE, Medical Imaging 2007.

Lee G. and Adrian E. Park. “Development of a More Robust Tool for Postural Stability Analysis of Laparoscopic Surgeons”. Surg Endosc (2008) 22:1087–1092

Lee G., T. Lee, D. Dexter, R. Klein, and A Park. “Methodological Infrastructure in Surgical Ergonomics: A Review of Tasks, Models, and Measurement Systems”. Surgical Innovation, Volume 14 Number 3, September 2007 153-167

Lee G., S. M. Kavic, I. M. George and AE. Park. “Postural Instability Does Not Necessarily Correlate to Poor Performance: Case in Point”. Received: 7 August 2006/Accepted: 22 September 2006/Online publication: 8 February 2007, Surg Endosc (2007) 21: 471–474

Lee G., T. Lee, D. Dexter, C. Godinez, N. Meenaghan, R.Catania and AE Park. “Ergonomic Risk Associated with Assisting in Minimally Invasive Surgery”. Accepted August 2008, Surgical Endoscopy

Appendices

A. Manuscript: Perioperative Scheduling Group

B. MVP: The Maryland Virtual Patient

C Novel, Web-Based, Information-Exploration Approach for Improving Operating Room Logistics and System Processes

D. A Research Portfolio for Innovation in the Surgical Environment

E. Development of a more robust tool for postural stability analysis of laparoscopic surgeons

F. Methodological Infrastructure inSurgical Ergonomics: A Review of Tasks, Models, and Measurement Systems

G. Postural instability does not necessarily correlate to poor performance: case in point

H. Ergonomic risk associated with assisting in minimally invasive surgery

Appendix A:

Manuscript: Perioperative Scheduling Group
Paul Nagy, PhD.

Abstract

Routine clinical information systems now have the ability to gather large amounts of data that surgical managers can access to create a seamless and proactive approach to streamlining operations and minimizing delays. The challenge lies in aggregating and displaying these data in an easily accessible format that provides useful, timely information on current operations. We describe a Web-based, graphical dashboard that can be used to interpret clinical operational data, allow managers to see trends in data, and help identify inefficiencies that were not apparent with more traditional, paper-based approaches. The dashboard provides a visual decision support tool that also assists managers in pinpointing problem areas in which the greatest benefits can be achieved by using business intelligence techniques to target time and energy toward continuous quality improvement. We review the limitations of paper-based techniques, the development of our automated display system, and key performance indicators in analyzing aggregate delays, time, specialties, and teamwork. Strengths, weaknesses, opportunities, and threats associated with implementing such a program in the perioperative environment are summarized. This research suggests Web-based tools can be made for targeted audiences and adjusted by role, position, or location with results in total participation in quality improvement and constant feedback that provide long-term rewards in cost efficiencies, staff and physician satisfaction, and improved patient outcomes.

Introduction

Management of the modern perioperative environment is a challenging act of balance and orchestration that often tilts perilously close to chaos. Many unforeseen delays (including but by no means limited to patient transport, case cart preparation, consent forms, and slow turnover) can trigger a cascade of events that escalate throughout the day, resulting in frustration for physicians, staff, and patients. The cumulative effects of many small and interacting delays keep the operating room (OR) from running at peak efficiency and can, in some cases, contribute to more serious errors in management and care.

Managers trying to address these delays in an ad hoc fashion find themselves playing “whack-a-mole” with serial problems: as soon as one glitch is resolved, another rears its head. The result is a reactive approach that focuses on immediate problems at the expense of the time and effort needed to identify root causes and long-term solutions that will prevent recurrences. The good news is that various routine clinical information systems now gather enormous volumes of data that surgical managers can access to create a seamless and proactive approach to streamlining operations and minimizing delays.

The process of leveraging these data in support of routine improvements, however, presents its own challenges, particularly in rethinking traditional reporting and analysis techniques. In the past, paper spreadsheet–based reporting methodologies have been used in management meetings, an approach that is insufficient to handle or analyze even the broadest trends in the increasingly large volumes of useful data collected. This

traditional, tactical approach most often focuses on only the short-term history of operations and fails to identify the small, recurrent delays that may occur across services.

Transparency and a broad scope of accountability are widely recognized as hallmarks of high reliability and dedication to quality in health care organizations.[1] These values, along with an emphasis on verifiable metrics and automated means of collection and assessment, have figured in significant advances in operations research in the management of the surgical environment in the past 5 years.[2-6] These advances contribute to the fulfillment of important goals of surgical units, including patient safety, access to ORs, economic efficiency, waiting time, and staff satisfaction.[6] Moreover, they have provided novel information about what factors contribute to which specific quality goals. Simulation studies using operations research, for example, have indicated that although immediate quality improvements in patient safety, waiting time, and satisfaction on the day of surgery should be a primary focus, only longer term decisions on staffing will provide economic efficiencies.[6,7] Thus reduction in turnover time in general will not result in increased volume,[8] but access to historical data and application of operations research methods can point to staffing solutions that will optimize economic efficiency.[9]

Our goal, as part of a grant on the OR of the Future from the Telemedicine and Advanced Technology Research Center, was to accelerate the adoption of these advances by providing an automated, holistic view of operations that would enable managers to discover patterns and causes of delays. We created a Web-based, graphical dashboard that could be used to interpret clinical operational data, allow managers to see trends in data, and help identify inefficiencies that were not apparent with more traditional approaches. This dashboard was designed to provide a visual decision support tool that

would also assist managers in pinpointing problem areas in which the greatest benefits could be achieved by applying time and energy toward continuous quality improvement.

What is Business Intelligence?

The field of business intelligence, sometimes referred to as business analytics, is the utilization of data warehousing, data mining, modeling, and forecasting to aid in managerial decision support systems.[10,11] Business intelligence is defined as "extracting useful information from the data generated by operational systems of an enterprise."[12] Many top corporate executives use business intelligence–generated electronic scorecarding and dashboarding methodologies to manage their operations with real-time decision making support. A 2007 Gartner, Inc. worldwide survey of 1,400 chief information officers ranked business intelligence as the number one technology priority for remaining strategically competitive.[13] Business intelligence methodologies extend directly to consumers in certain markets. Financial Web sites provide individual investors with extensive research and graphical performance analysis of publicly traded companies.

Large academic medical centers, which often generate revenues in excess of $200 million, are in the same financial league as the medium-to-large businesses in which dashboards are commonplace. Yet few medical centers have invested in the development and routine implementation of tools to analyze and improve the efficiency and effectiveness of perioperative management and operations. Many ORs continue to "fly blind" with regard to concepts such as indexing and performance measurement.

Internal graphical dashboards have been proven in other environments to provide useful and productive platforms for continuous quality improvement. The Six Sigma quality methodology, for example, frequently employs dashboards for process

management.[14] A dashboard can provide a consistent framework of defined metrics, known as key performance indicators, that aid in defining and redefining quality and goals as well as offering quantifiable data on achievements.[15] Evidence of consistent improvements through a public dashboard is then used to help align the various parts of an organization to target enhanced performance.

The Potential of Information Visualization

Our project was designed to provide a visual knowledge exploration system to assist managers and senior leadership in understanding trends and patterns. The tools employed in this system provide interactive views of data at various granularities and in a series of graphical or tabular formats. These tools can quickly and with minimal user effort impose various types of analyses on the full dataset or on interactively selected subsets of data.

The goal of a visual knowledge exploration system is to provide tools that facilitate interaction with information in an easy, transparent, and meaningful manner. A well designed graph can tap into the pattern-recognition capabilities of the human visual system. In certain types of patterns, human vision can identify a unique (outlier) value within 200 msec, regardless of whether few or many data points are present.[16,17] However, this ability is entirely dependent on the manner in which the pattern is displayed. Proper visual display is crucial to the use of large datasets for complex decision making support. The optimal type of data display has been the focus of a substantial body of literature and reporting, and the definitive answer changes as rapidly as new technologies enter the information arena.[18-22] Some studies suggest the superiority of graphical formats (bar charts, pie charts, etc.) over tabular presentation (data tables) for

certain tasks, whereas the reverse is true for other tasks. More recent work indicates that a constellation of factors must be considered in determining the most advantageous dataset display formats, including type of task, underlying structure of the data, and the knowledge level of the users.[23]

What's the Problem with Paper-Based Reporting?

The benefits of graphical dashboarding can be appreciated more fully by looking at the limitations of traditional paper-based reporting in identifying and managing ongoing operations challenges. Understanding these limitations is important, because in many institutions paper-based reporting is so engrained into routine practice that clinicians and perioperative managers may find it difficult to take the steps needed to adapt to other methodologies. Paper-based reporting management systems are limited in the following areas:

(1) *Time.* Significant time is required to gather information from various sources and compile reports by hand. Decision making is a time-sensitive activity that requires actionable information. Decision making, a process that that should be based on "fresh" data, is adversely affected when time simply does not permit preparation of all possible permutations of analyses that might be informative and useful.

(2) *Effort.* The inherent limitations of paper restrict the number of questions than can be asked and tend to generalize rather than drill down in areas of analyses. Expanding the scope of a paper report requires extra labor. Most often, the result is a trade-off between the time required to generate the report and the quality of effort required in preparing the results for analysis.

(3) *Hindsight.* One of the most frustrating characteristics of paper-based reports is that they provide only answers to questions that were identified *before* the management meeting and discussion. New questions asked during the meeting must be tabled until the next meeting so that analysts can gather the new information required. These tabled questions prevent more purposeful discussions about the data and leave managers with limited information to support decision making in the short term.

(4) *Scope of report.* The amount of information that can be contained in a paper-based report is limited, as is the amount of information that can be reviewed within a reasonable amount of time. Selectivity becomes a necessity; yet it is difficult to predict which questions managers will have during any given meeting. Attempts to broaden the scope of paper-based reports can be both time consuming and problematic: the larger the amount of data in a paper report, the more difficult it is to find any specific piece of information.

(5) *Granularity.* Aggregate statistics do not allow the user to drill down to understand the underlying distributions to evaluate credibility. Mean statistics offered in most paper-based reports are unreliable when describing non-normal distributions of data. A single chart or table on paper can show only one view, and it is difficult to present both overviews and detailed information in a single presentation. Showing trends can obscure source data, where showing only source data can obscure trends. Including both or all in a paper-based report can be labor/time intensive to review and is impractical as a routine practice.

(6) *Multiple versions of truth.* Different groups generating separate reports or even the same reports at different times can result in conflicting operation directives that can

add to confusion in effective decision making. The human intervention inherent in paper-based reporting can also introduce bias that may lead to data analysis errors. Moreover, the passage of time between the collection/analysis of data and final reporting may mean that no direct links exist between current operational data and the paper-based report.

Dashboard Design

In developing the toolkit and dashboard described in this article, we followed the set of principles for effective information seeking outlined by Scheiderman.[24] Effective systems should provide an initial graphical overview of the data, allow the users to zoom in on and filter data in the overview, and then receive details about specific data points on demand. In order to create a our concise visualization environment, we created graphs that were “clickable” within a drill-down interface to provide fast and intuitive zooming and filtering of data. These graphs also provide detailed information about data points when the user hovers over a data marker with the mouse. One of the core requirements for a management dashboard is that it be Web based to allow secure access to all authorized users from any location at any time.

Standard data warehousing techniques were used extraction, transformation, and loading. The database was populated by parsing a text file in a comma-separated delimiter format provided by Cerner SurgiNet Surgical Information System (Kansas City, MO). The text file was converted to ANSI Structured Query Language commands as inserts using the MySQL database administration utility PhpMyAdmin. Data were anonymized for patient information, because the focus of the system was operational efficiency, not identifying specific patient-associated incidents. Six months of operational data were uploaded into the system, incorporating performance statistics on 7,807 cases

on 8 MB of disk space on the server. These cases incorporated all of the operating rooms with both inpatient and outpatient admissions.

Identifying Key Performance Indicators

The American Association of Clinical Directors derived a common glossary of the exact meaning of times used for scheduling and monitoring surgical procedures.[25] Time stamps were extracted from the clinical database for: scheduled start time, time at which the patient enters the operating room (PIR), the time at which surgery begins (PST), surgery end time (PF), time at which the patient leaves the room (POR), and the turnover time of the room (TOT). As a hospital policy, when a case begins 15 or more minutes later than scheduled, the circulation nurse must specify a reason for the delay. These performance data are combined with data from the case such as the room, surgeon, anesthesiologist, case number of the day, and the service.

Organization of the Web Site

The Web site was designed to allow analysis from several perspectives. One of the principle mantras of information visualization and data discovery, identified by Schneiderman, is the ability to "overview first, zoom, filter, then details-on-demand." [26] The ability to view data from multiple perspectives assists and increases confidence in decision making. The user accesses each category via a navigation bar of tabs at the top of the web site. As the user navigates through the system, a trail (called a "breadcrumb" and shown in Figure 1) is displayed to illustrate how the user navigated to that point and to allow easy backtracking.

Aggregate Delay Analysis

A total of 43 delay types were identified as reasons for delays. These were grouped into general root causes of materials, patient, prerequisite task(s), scheduling, staff, and transport. At the top of the delay analysis page, as shown in Figure 1, pie charts demonstrate the relative number of delays per root cause and their cumulative impact in time. Although some delays are not numerous, they might have a large effect on the operations of an OR. The user clicks on the pie chart, selects a root cause, and is presented with an analysis page of all the underlying delay causes for that root cause, broken down in the same way by relative number and impact. By selecting a delay cause, the system moves to show a breakdown by specialty, displaying the number of incidents and their average delay times. Selecting a service displays all the cases, and by selecting a case the details for that case can be displayed. Within the span of 4 clicks, a user can drill down from all the cases in the database to the details of an individual case. The delay analysis tool is useful in understanding the cumulative cost of systemic delays and which specialties are most affected by them.

Temporal Analysis

The temporal perspective provides a daily tactical review of cases to determine over a specified period of time which ones were delayed and why. To present the utilization levels of the ORs in a given day, we used a polar chart showing cumulative room utilization as a function of the hour of the day (Fig. 2). This is useful for look at the relationship between room utilization and staffing levels. The user can drill down to the specifics of a single case or choose to look at data grouped by room or specialty. System delay types, such as transport issues, can affect multiple rooms and specialties across suites of ORs over different periods of time.

Service Analysis

Service analysis focuses on key performance indicators within each specialty. Medical specialties within the OR have widely differing dynamics for case efficiency, utilization, turnover, and case length based on a number of factors, including but not limited to procedure complexity and patient acuity. For some types of data analysis involving services or subspecialties, bubble charts provided a useful way to organize data (Fig. 3). The bubble chart plots each service by its average case length in the *x* axis and average delay duration in the *y* axis. The size of the bubble for each specialty is directly proportional to the number of cases performed. The more cases a service performs, the larger the diameter of the bubble.

For each specialty analysis, the site provides histograms for case length and delay duration. Histographic analysis is useful in determining the distribution type, the spread of the distribution, and the existence of outliers that may distort statistical analysis.

Another display generated was a scatter graph of all cases plotted by their scheduled case lengths compared with actual duration (Fig. 4). Regression analysis shows potential correlations, along with graphical bands illustrating confidence intervals for the line and the points. This index of predictability of scheduling is especially useful in identifying and drilling down on the outlier cases to understand their causes of variance. Each diamond represents an individual case, and, by clicking on a diamond, the details of that case are displayed (Fig. 5).

Teamwork Analysis

With a surgeon and principal anesthesiologist assigned to each case, we can display delay causes for those cases. As shown in the spider graph in Figure 6, delays are

grouped by and aggregated by the blue bars. The farther out the bars, the greater were the number of occurrences for that root cause. In an overlapping orange are the average delays by root cause for all physicians in that specialty, normalized by the number of procedures done by that physician.

Using this teamwork analysis, it is possible to identify specific teams that appear to work well together as well as those that are not routinely time efficient. Other factors, of course, must be considered in reviewing these data, and it would be difficult to determine root causes for efficiency or inefficiency in a specific case. However, this knowledge may provide strategic information that could contribute to what business intelligence experts call a "discovery cascade."

Results

Results of an initial rollout of the Web system were assessed through interviews with senior management. This included discussions with the chief medical officer, chief operations officer, chief nursing officer, chairs of surgery and anesthesiology, and several perioperative managers. In presenting this potentially disruptive tool to management, we performed a strengths, weaknesses, opportunities, and threats analysis (known in business intelligence parlance as a SWOT analysis) to classify their observations.

Strengths

Our Web-based approach was seen as a powerful tool that would aid management in identifying systemic, process-driven root causes for delays and other problems and that had the potential for positive effects on the culture of the organization. Among the positive aspects they cited were: (1) This approach turns traditional paper-based data into knowledge and presents this knowledge in easy to digest chunks. (2) The dashboard is

independent of any single vendor. (3) If used with a data repository, it has the potential ability to link data from different information systems. (4) It provides a systemic view that can calculate the total costs of root causes. (5) It provides a quick visual way to target improvement. (6) Additional metrics can be added at the request of management with minimum programming effort. (7) Visual displays made outliers and trends more easily identifiable and rendered distributions more easily understood than standard aggregate statistics.

Weaknesses

The reviewers identified 4 areas of weakness and potential improvement for the Web-based system. (1) Timeliness: Depending on the method of data extraction, data may not be live or near-live. If data are provided via an upload, they will be only as recent as the last event. The dashboard optimally should have an Open Database Connectivity connection to a clinical data repository (CDR) or similar copy of the live production environment. The update schedule of the CDR will determine the timeliness of the dashboard.

(2) Personnel resources: Skilled personnel are required to build and maintain a graphical dashboarding application. A surgical informaticist is a good choice as they have the clinical domain experience combined with the principles of management and information technology. This person needs to guide the development of metrics by using clinical knowledge to extract meaningful and relevant data. This individual can also play a crucial role in bridging the cultures among health care providers, information technology specialists, and business process managers.[27]

(3) Hardware resources: Hardware resources include access to server space with sufficient processing power and storage to handle a large database. The database storage space required is minimal; however, the central processing unit that drives the data mining must be powerful.

(4) Management training: The introduction of analytics and acceptance of business intelligence practices within a group, particularly one that already has a long-engrained operations process, cannot be accomplished overnight.[11] An investment of time and effort is required and involves education of managers on the use of these tools and the ways in which they can be incorporated into decision making. In the process, the focus of the managers and the entire organization should change from trying to understand the latest event to looking at trends within the data to predict what will happen next and identify ways to achieve the best possible results.

Opportunities

A dynamic surgical block utilization chart with easily available drilldown, as created in our project, allows surgical chiefs to continually monitor utilization. The drilldown permits them to see which days are being underutilized and by whom and points to immediate courses of action rather than waiting for end-of-year retrospective and analysis. When competition is fierce for OR time, this transparency can be extraordinarily valuable for surgical practices.

Another potential benefit that can accrue from matching staffing with caseload to optimize OR efficiency.[28] Cases in overutilized time are 1.75 times more expensive than cases during normal staffing hours[29], the goal is to match case load with staff.[29] The dashboard tool pulls in scheduling data from the clinical information system and can

display the number of projected cases at 1-hour intervals. The dashboard can also use retrospective data on add-on cases to estimate a caseload probability by the hour. To maximize efficiency, a user input can be created whereby a manager or charge nurse may enter data on staffing levels, helping to match case load to staffing.

Along with scheduling efficiency, OR senior staff and managers may want to match clinical proficiencies with cases. Displays can be created to show circulator/scrub combinations with surgical specialties and case types similar to the surgeon−anesthesia graphs presented. This gives managers opportunities to maximize good teams and identify teams that need improvement.

Many opportunities are available for benchmarking between services and between organizations. The only limitation is the ability to capture data from an information system or network, apply a meaningful analysis, and provide an easily understood graphic for the appropriate audience.

Current Procedure Terminology (CPT) codes would be an important additional piece of information, because case efficiency should be benchmarked against similar cases. Cardiac Thoraic cases, for example, have long turnover times because of the degree of complexity involved in setup of equipment, drawing of drugs, patient preparation, etc. National benchmarking can be imported to compare against the organization's benchmarks reports for CPT and DRGs, morbidity and mortality, length of stay, and complications and presented in an easy-to-navigate and -understand visual.

Opportunities for assessing clinical outcomes include measures such as infection control, preoperative antibiotic compliance, unplanned returns to the OR, staff compliance on chart quality, timeliness, completeness, staff arrival time, etc. Financial

reports can include direct costs, indirect costs, contribution margins per case/specialty, labor costs, supply costs per case, and metrics associated with defining and monitoring best practices.

Threats

Two potential threats to a system such as the one we devised were identified by the interview group and by our own developers.

The first threat is in the area of data quality and integrity. The data retrieved for our clinical information system have 2 sources of origin. First, scheduling data are obtained. These include but are not limited to scheduled start date and time, duration, procedure, surgeon, and anesthesiologist. The schedulers at our institution reside both centrally in a surgical posting office and decentralized in physicians' offices (in the oral maxillofacial and organ transplant services). Because scheduling data do not directly enter the patient's medical record or roll directly into clinical documentation that must be reviewed and modified by a nurse, it is assumed that the risk for bias is minimal. Manipulation of case durations and scheduled start times is limited by system controls. The surgical posting office does have the ability to override system data (e.g., for scheduled case duration, which is a byproduct of historical averages), but this is not done without approval from a supervisor.

Data from nursing documentation is under constant review by various clinicians to audit work. This process ensures data integrity and compliance and serves as a modest check and balance. Most of these data are objective, and although some bias may be present, this will most likely be minimal. The area of documentation that is most prone to bias is the "delay reason," because of its highly subjective nature and possible

repercussions from management. Another factor for inaccurate delay reporting is the phenomenon of cascading delays (i.e., when a delay early in the day causes delays in subsequent cases). By the time of the third or fourth delayed case, it is difficult to ascertain the cause other than to note that the previous case "ran over." During this series of delays, an entirely different cause of delay may happen in a specific case, but the reference point of a scheduled start time is lost, so that it is much more difficult to document a delay cause and duration. Time stamps (in-room time minus scheduled start time), of course, provide well documented and precise record of delay in minutes, but this does not qualify delay by reason type and provides no insights for root cause analysis.

The second threat to initiation of a system such as the one we developed lies in the general perceptions by staff and physicians. Many may feel that they are being spied upon or monitored, especially in areas in which no previous metrics existed. Others may find themselves out of their routine comfort zones. Underperforming staff who worry that they may be identified by the system may aggressively resist implementation of the new tools or work to undermine data integrity. Depending on the organization's structure, the open availability of data could result in "punishment" for individuals or a group rather than the intended promotion of positive departmental and institutional change. Moreover, in environments in which competition for OR time is strong, surgical chiefs may be tempted to use data as a weapon to promote their own agendas.

The transparency of the data should alleviate some of these concerns. Team members should be able to see the data in which performance is being judged. In the past, data was obtained by someone walking around with a clipboard or in a back office recording data off charts, with limited or no ways to verify whether or the data were true

and accurate. With the drilldown features and different ways of organizing data for dashboard display, team members can easily view the raw data.

Conclusion

Strategic decisions made on the basis of management instinct have a lasting impact on the well-being of an organization. Management could benefit from the adoption of business intelligence tools that provide a quantifiable, validated alternative to instinct and ad hoc choices decision making.

Behavior and practice changes are central to achieving the objective of quality reports that drive efficiency. Too often data are not integrated within the scope of daily practice. Acceptance of the importance of data must become a part of the culture of the organization. Graphical dashboards that present information down to the simplest, easiest-to-understand, and most accurate levels can compel this behavior change. Managers in the perioperative environment should seize the opportunity to integrate data into their organizations' cultures. Our research suggests that one promising approach is in Web-based tools that can be made for targeted audiences and adjusted by role, position, or location. The result can be total participation in quality improvement and constant feedback that provides long-term rewards in cost efficiencies, staff and physician satisfaction, and improved patient outcomes.

Acknowledgements

This project was supported under an OR of the Future grant from the Telemedicine and Advanced Technology Research Center. I would like to thank Dr. Nancy Knight from the University of Maryland for her expert assistance in preparing this manuscript.

References

1. Weick KE, Sutcliffe KM: Managing the Unexpected: Assuring High Performance in an Age of Complexity. San Francisco, CA: Jossey-Bass, 2001.
2. Dexter F, Xiao Y, Dow AJ, Strader MM, Ho D, Wachtel RE: Coordination of appointments for anesthesia care outside of operating rooms using an enterprise-wide scheduling system. Anesth Analg 105:1701-1710, 2007.
3. Dexter F: Why calculating PACU staffing is so hard and why/how operations research specialists can help. J Perianesth Nurs 22:357-359, 2007.
4. Dexter F: Bed management displays to optimize patient flow from the OR to the PACU. J Perianesth Nurs 22:218-219, J2007.
5. Dexter F: Operating room utilization: information management systems. Curr Opin Anaesthesiol 16:619-622, 2003.
6. Dexter F, Epstein RH, Traub RD, Xiao Y: Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. Anesthesiology 101:1444-1453, 2004.
7. Macario A, Chow JL, Dexter F: A Markov computer simulation model of the economics of neuromuscular blockade in patients with acute respiratory distress syndrome. BMC Med Inform Decis Mak 6:15, 2006.
8. O'Sullivan CT, Dexter F, Lubarsky DA, Vigoda MM: Evidence-based management assessment of return on investment from anesthesia information management systems. AANA J 75: 43-48, 2007.

9. O'Neill L, Dexter F: Tactical increases in operating room block time based on financial data and market growth estimates from data envelopment analysis. Anesth Analg 104: 355-368, 2007.

10. Davenport TH: Competing on analytics. Harvard Bus Rev. January 2006:1-12.

11. Davenport TH, Harris JG: Competing on analytics: the new science of winning. Boston, MA: Harvard Business School Press, 2007.

12. Chisholm M: The twin towers of BI babel: enterprise architecture. BI Rev. December 2007. Available at: www.bireview.com/issues/2007_42/10000440-1.html. Accessed: January 10, 2008.

13. Beer S: Business intelligence top priority of CIOs. itWire. February 2007. Available at: www.itwire.com.au/content/view/9906/53/. Accessed: January 10, 2008.

14. Pande PS, Neuman RP, Cavanagh RR: The Six Sigma Way: Team Fieldbook. New York, NY: McGraw-Hill, 2002.

15. Malik S: Enterprise Dashboards: Design and Best Practices for IT. Hoboken, NJ: John Wiley & Sons, 2005.

16. Treisman A: Preattentive processing in vision. Comput Vis Graphics Image Processing 31:156-177, 1985.

17. Treisman A, Gormican S: Feature analysis in early vision: Evidence from search asymmetries. Psychol Rev 95:15-48, 1988.

18. Washburne JN: An experimental study of various graphic, tabular, and textual methods of presenting quantitative material. J Educ Psychol 18:361-376, 465-476, 1927.

19. Tufte ER: The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press, 1983.
20. Cleveland WS, McGill R: Graphical perception and graphical methods for analyzing scientific data. Science 229:828-833, 1985.
21. Montazemi AR, Wang S: The effects of modes in information presentation on decision-making: a review and meta-analysis. J Manag Inform Syst 5:101-127, 1988.
22. Feldman-Stewart D, Brundae MD, Zotov V: Further insight into the perception of quantitative information: Judgments of gist in treatment decisions. Med Decis Making 27:34-43, 2007.
23. Meyer J, Shamo MK, Gopher D: Information structure and the relative efficacy of tables and graphs. Hum Factors 41:570-587, 1999.
24. Card SK, Mackinlay JD, Shneiderman B: Readings in information visualization: using vision to think. San Francisco, CA: Morgan Kaufmann Inc., 1999.
25. Procedural Times Glossary of the AACD. AACD. October 2005. Available at: aacdhq.org/Glossary.htm. Accessed: January 10, 2008.
26. Shneiderman B: Inventing discovery tools: Combining information visualization with data mining. Informat Visualiz 1:5-12, 2002.
27. Charters KG: Nursing informatics, outcomes, and quality improvement. AACN Clin Issues 14:282-294, 2003.
28. Dexter F, Ledolter J, Wachtel RE: Tactical decision making for selective expansion of operating room resources incorporating financial criteria and

uncertainty in sub-specialties' future workloads. Anesth Analg 100:1425-1432, 2005.

29. Strum DP, Vargas LG, May J: Surgical subspecialty block utilization and capacity planning: A minimal cost analysis model. Anesthesiology 90:1176-1185, 1999.

Appendix B:

MVP: The Maryland Virtual Patient

Project Report

Dr. Bruce Jarrell
Dr. Sergei Nirenburg
Dr. Marjorie McShane
Dr. Stephen Beale
Dr. George Fantry

February 28, 2008

Summary

In Year 4 of the project, our team has delivered two new versions of the Maryland Virtual Patient Environment. The realism of the simulation has been enhanced by including coverage of "unexpected" interventions; allowing discontinued treatments; allowing new diseases to develop due to side effects of treatments. The user interface has been redesigned. A new agent-based architecture has been developed to support enhanced cognitive capabilities of the virtual patient and the intelligent tutor, including language capabilities. In the area of language processing, a dialog processing model was developed. Work has continued on improving the language understanding capabilities, centrally including treatment of referring expressions. Enhancement of static knowledge resources, the ontology and the lexicon, has been ongoing. Work on extending the coverage of diseases has been ongoing: a further improvement of the model of GERD is under way, as is the modeling of cardiovascular diseases. A totally reworked system version, with dialog support, is planned for release in June 2008. Work has also been ongoing on improving and extending the set of development tools – the DEKADE demonstration, evaluation and knowledge acquisition environment supporting natural language work has been revamped; the interface for creating instances of virtual patients has also been enhanced; a web-based environment for supporting internal documentation has been installed. Finally, we have written, submitted, published or delivered X conference and journal papers.
This report introduces the basics of the MVP approach, discusses its place on the map of intelligent systems in clinical medicine and describes the project's status and research and development activity presently under way.

Background

Many medical educators believe that the current system of medical education in the US fails to reliably provide students with a sufficient breadth of clinical experience to ensure the development of clinical diagnosis and treatment skills. Students typically manage too few patients with too few clinically relevant variations of a disease to become first-rate care providers by the time they complete their residency. In addition, learning clinical medicine through experience on live patients imposes a heavy responsibility, offering no channel for learning by trial and error. One way to circumvent these shortcomings of medical education is through interactive computer systems that simulate a clinical care environment.

The specific goal of the MVP project is to create a computer system for teaching medical students cognitive skills of an attending physician related to diagnosing and treating patients. The system is intended to provide a safe, hands-on environment in which the students communicate with and treat simulated virtual patients as well as obtain help from an automatic mentor. The system environment also approximates the real world in making the student a member of a team of medical professionals – simulated lab technicians and specialist consultants.

For a system of this kind to be successful, it must be **realistic** in a variety of ways. Diseases and their progression must be realistically simulated, using biomechanistic knowledge whenever possible but reverting to expert clinical knowledge in situations where biological mechanisms are still unknown. Virtual patients must be capable of carrying more than one disease at a time. Side effects of treatments must be modeled to allow dynamic changes in the pathophysiological states of virtual patients. The above implies that virtual patients cannot be fully pre-scripted, which means that simulations must be based on causal modeling of event sequences. Virtual patients must be able to model symptom perception, communicate with the physician and make reasoned decisions relating to treatment. Mentoring knowledge should reflect the best clinical practices in diagnosis and treatment. A broad inventory of diseases must be covered. Finally, even for a single disease, a pedagogically sufficient number of virtual patients must be created (or allowed to be created) with different genetic predispositions, disease progressions and reactions to treatment.

To meet the above requirements, a computer system must have several **enabling capabilities**. The simulation module must be able to deal with knowledge of different provenance and varying grain size, on a scale from, e.g., well-understood cell-level biochemical mechanisms to, e.g., qualitative expert experience with treatment outcomes or co-morbidities. The virtual patient must be capable of perception, reasoning and action. This makes it a classical example of an artificial cognitive agent. Language is the most natural way of communicating with the human user of the system and the only truly realistic one. Finally, the amount and variety of medical knowledge required for building a realistic system is non-negligible.

To maintain realism, the system must be meaningfully responsive in novel, unexpected, unscripted situations. The main source of this novelty is unpredictability of the actions of the human (the medical student). There are two ways in which the human can introduce novel situations – through treatment interventions and in language-based communication with the virtual patient and the tutor.

For a system of this nature to be **feasible**, a number of constraints must be introduced in its design. It is clear that the coverage of medical knowledge must be incremental, starting with a subset of diseases and gradually increasing their inventory. The grain size of disease description can also start with a minimally useful level of coarseness and made progressively finer over time in successive versions of the system. The simulation of auxiliary agents – lab technicians and specialist consultants – can be at first skeletal. It is possible initially to exclude direct visual and haptic physical examination of the patient. The language communication can be in written form, not speech-based. However, the virtual patient and the tutor must be able to understand the content of the communication and the intentions of the human user encoded in written messages and be able to reason about how to respond appropriately. If communication is to be realistic, it will not be possible to have this understanding rely on a predetermined list of expected questions with their answers. Similarly, while it is possible to limit the set of intentions, goals, plans, character traits and action types modeled in the virtual patient, the overall mechanism of manipulating them must be developed. This is because neither pre-scripting the virtual patient's actions in response to a variety of human interventions nor determining what to do stochastically can guarantee sufficient realism in system behavior.

Why do we believe that our work is feasible?

First, our development efforts are targeted toward **specific applications**: there is no attempt to develop a fully generalized, plug-in ready cognitive architecture (like TRAINS/TRIPS), or to implement a broad-coverage, domain-independent dialogue system, or to equip system agents with all of the plans and goals of human beings, or to endow them with the full spectrum of possible character traits (as is done in theoretical approaches to affective modeling), or to model diseases at a grain size any finer than that needed to support the given application. Instead, theoretical and practical advancements are geared toward the near- and long-term future of the specific systems, with infrastructure decisions being made with a long-term view but knowledge support targeted at near-term goals.

Second, the **integrated approach to knowledge modeling** in MVP permits the same ontological substrate to be used for knowledge-based simulation, planning, and NLP, meaning that once knowledge is encoded it is available to all system agents and processors. The OntoSem ontology used in MVP already includes over 9,000 concepts, described by an average of 16 properties each; around 7,500 of those are from the general domain, with the remaining 1,500 devoted to medicine. Moreover, since scripts describing complex physiological and cognitive events are formally part of the ontology,

the same scripting language used for physiological simulation (which is already understood by our simulator engine) can be used for planning and dialogue.

Third, the dialogue processing model is grounded in the **OntoSem deep semantic natural language processing system,** which has been under development for over 20 years.

Fourth, the past decade has produced a valuable **body of research** on cognitive engineering, agent networks, planning, plan- and goal-centered dialogue systems, etc. This large body of work includes inventories of needs for intelligent systems, sample architectures, descriptions of problems encountered, bridges between descriptive, theoretical and implementational work, and reports from the field that provide a good understanding of the current state of the art. In short, this body of work is permitting us to quickly reap the benefits of hard-won insights.

The MVP project involves a broad variety of specific tasks. Knowledge about a) human physiology and pathophysiology, b) best clinical practices and c) specific instances of virtual patients must be acquired from experts. This knowledge must be represented in a way that facilitates automatic processing by computer. Methods of computer processing of this data must be developed, including a) disease progression simulation, b) reasoning capabilities to simulate the virtual patient's and virtual tutor's decision making, and c) communication capabilities between people and artificial agents. Finally, issues relating to a) software system architecture and maintenance; b) static knowledge resource maintenance; and c) testing and evaluation support must be addressed.

In what follows, we will describe the current status and immediate plans of the MVP project.

The Knowledge and Processing Architecture

Current Functionality of MVP

We present here a simplified, coarse-grained sketch of the MVP simulation, interaction and tutoring system (further details will be supplied later). A virtual patient instance is launched and starts its simulated life, with one or more diseases progressing. When the virtual patient develops a certain level of symptoms, it presents to the attending physician, the system's user.[1] The user can carry out, in an order of his or her choice, a range of actions: interview the patient, order diagnostic tests, order treatments, and schedule the patient for follow-up visits. The patient can also automatically initiate follow-up visits if its symptoms reach a certain level before a scheduled follow-up. This patient-physician interaction can continue as long as the patient "lives."

[1] Human users of the MVP system can be of various profiles, including medical students, residents, physicians or other medical personnel seeking to refresh or improve their skills in some area, physicians (both developers and non-developers) testing the system for accuracy and robustness, and examinees. Throughout this paper this heterogeneous group of users will typically be referred to simply as "users".

As of the time of writing, the implemented MVP system includes a realization of all of the above functionalities, though a number of means of realization are temporary placeholders for more sophisticated solutions, currently under development.[2] The most obvious of the temporary solutions is the use of menu-based patient-user interaction instead of natural language interaction. While this compromise is somewhat unnatural for our group, which has spent the past 20 years working on knowledge-based NLP, it has proved useful in permitting us to focus attention on the non-trivial core modeling and simulation issues that form the backbone of the MVP system.

MVP currently covers six esophageal diseases pertinent to clinical medicine: achalasia, gastroesophageal reflux disease (GERD), laryngopharyngeal extraesophageal reflux disease (LERD), LERD-GERD (a combination of LERD and GERD), scleroderma esophagus and Zenker's diverticulum.[3]

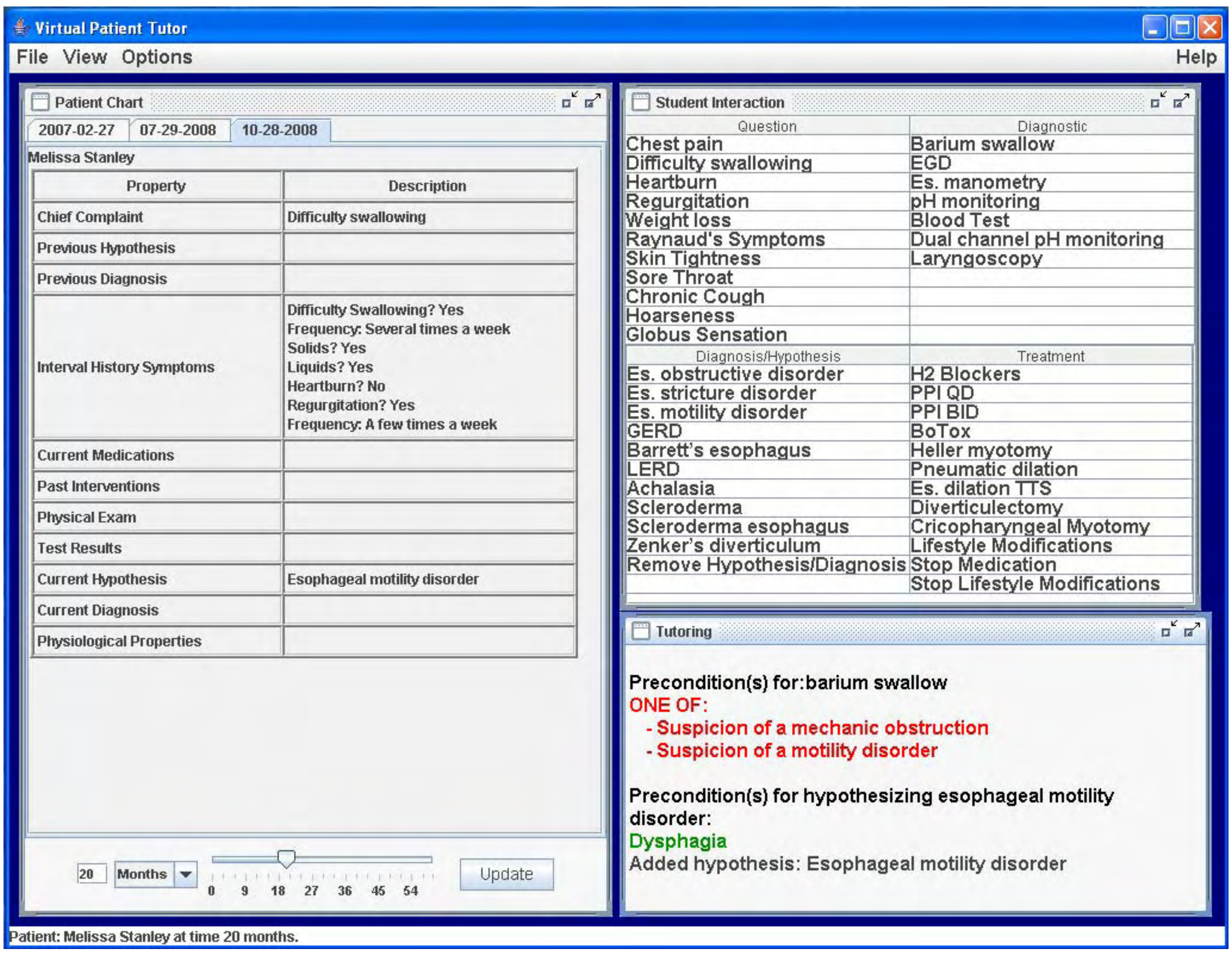


Figure 1. The main user interaction screen.

[2] The motivation for the use of the placeholders was the desire for rapid integration of a working prototype of the system. This goal was successfully attained, and at the time of writing the MVP has started to be used in the curriculum of the University of Maryland School of Medicine.

[3] Detailed descriptions of the diseases achalasia and GERD are given in McShane et al. *forthcoming* and McShane et al. *submitted*.

Figure 1 shows a screen shot of the main pane of the simulation and mentoring application. The left-hand side is the patient chart, with the page for the current visit presented. Records of previous visits can be reached using their associated dated tabs. The chart includes both information that is automatically generated from the previous visits and information that is compiled based upon the events that occurred during the latest visit. The information carried over from previous visits includes the chief complaint, the (previous) disease hypothesis and diagnosis, currently prescribed medications and past interventions. The information that is added based on the latest visit includes current symptoms (called Interval History Symptoms), results of the (simulated) physical exam, results of tests ordered during the given visit, and a disease hypothesis and/or diagnosis added during this visit. Depending on the specific tutoring setting, the user can be given access to a complete list of the virtual patient's physiological properties (the omniscient view of the patient).

The lower left pane contains a time slider, which permits the user to advance patient time at intervals of days, weeks or months. Other configurations of the system include other methods of managing simulation time.
The upper right quadrant contains four panes of menus with which the user can ask the patient relevant questions about symptoms, order diagnostic tests, posit a hypothesis and diagnosis, and order treatments.

- Each **symptom** button asks all relevant questions about the given symptom at once: e.g., asking about heartburn queries the patient about whether he/she ever has heartburn and, if so, follows up with questions about frequency, severity, and whether any kinds of food cause heartburn.[4]
- Each **diagnostic** button orders that diagnostic test, with the results appearing as fillers of Test Results in the patient chart.
- Each **diagnosis/hypothesis** button permits the user to select that disease as a diagnosis or hypothesis, with the associated slots of the patient chart being filled accordingly.[5] For system configurations that include automatic tutoring, the recording of hypotheses and diagnoses – which make explicit the user's current thinking – is needed for the evaluation of whether a move is being made for the right reason.
- Each **treatment** button carries out a treatment, with the treatment then recorded in the Past Interventions cell of the current and subsequent patient charts.

The lower right quadrant contains mentoring messages for configurations when the tutor agent is enabled. (See section 2.7 for further discussion.)

Returning to the patient chart, it can provide access to the omniscient view of the patient's physiological properties, a feature that can be enabled and disabled under various system configurations. For example, Figure 2 shows a filtered subset of the

[4] When physicians and students tested the system, the preference for asking nests of questions with a single click was overwhelmingly confirmed. When natural language support is added, questions will be asked in series, mimicking office interviews.
[5] After the initial choice of a disease from a menu, a pop-up box asks whether this is a diagnosis or a hypothesis.

properties of Melissa Stanley 20 months into her disease. The values in red are those that changed since the last time the clock was advanced. This physiological view of the virtual patient is useful not only for validating system functionality, it also has pedagogical uses: a user can, under certain configurations that the instructor can control, watch property values change in response to various interventions. This physiological view drives home the point that the MVP system is not an inventory of stored scenarios: it is a live simulation whose outcomes are crucially dependent upon user actions.

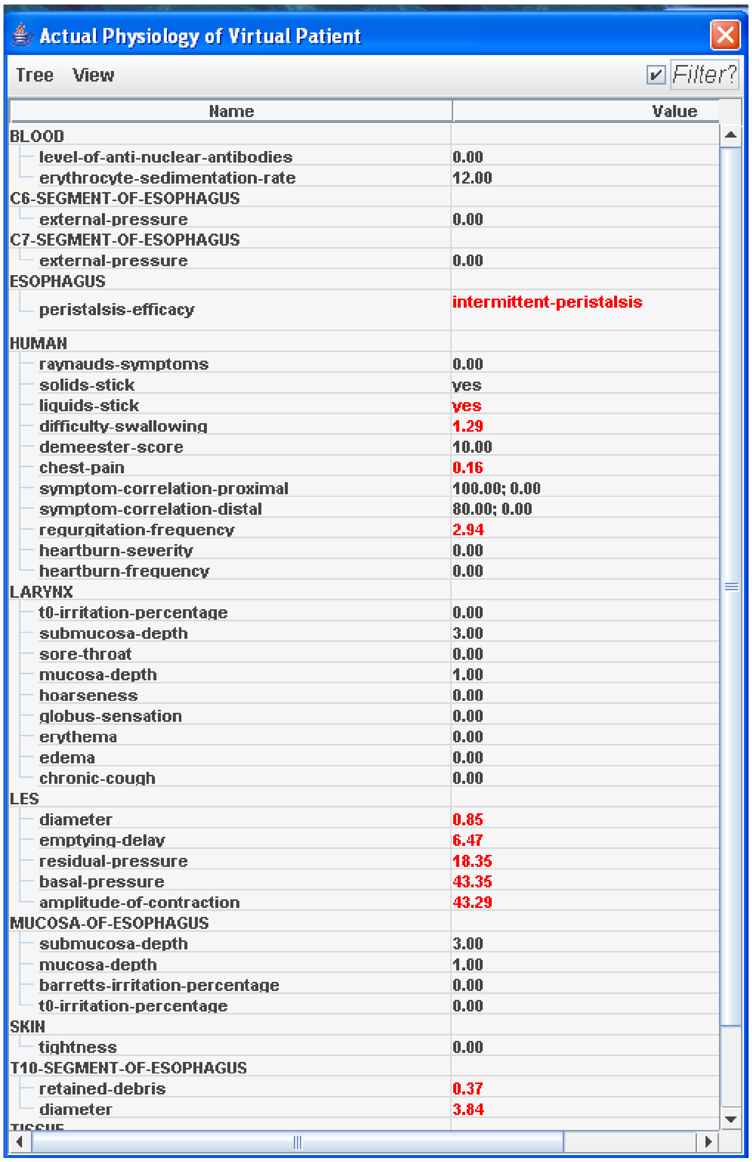


Figure 2. A subset of “Melissa Stanley”’s property values at month 20.

An Example of Patient Management

To illustrate the system operation, we present below an example of how a user might handle the case of “Melissa Stanley,” the patient a subset of whose profile is shown in Figures 1 and 2. The sequence of steps in this example represents only one of thousands

of possible paths through this patient's case. Indeed, different users, working under different tutoring intensity settings, will create very different system runs. At the start of this particular run the tutor agent is turned on to the maximal information-providing tutoring setting. With this setting, the tutor agent reveals, at appropriate times, all fulfilled and unfulfilled preconditions (marked in green and red, respectively) for user actions, and user actions are only permitted if the necessary preconditions are fulfilled. In this illustration, the agent of the action at each step is *italicized*, and the action itself is marked in **boldface**.

1. The *VP instance*, Ms. Stanley, **presents** with the chief complaint "difficulty swallowing". The time of presentation – 17 months into the disease – is indicated by the position the time slider but can be hidden if desired for pedagogical reasons.
2. The *user* **asks** about heartburn, difficulty swallowing and regurgitation.
3. The *VP instance* (Ms. Stanley) **responds** stating that the only available positive symptom is mild difficulty swallowing occurring less than once a week (the patient has the personality trait of being a hypochondriac, meaning that she presents to the doctor given very mild symptoms).
4. The *user* decides that the symptoms are not severe enough to be acted upon and **schedules** a follow-up visit in 3 months. This ability to monitor a patient over time, with or without intervention, is a crucial aspect of clinical medicine not targeted by any other simulation systems.
5. The clock advances to 20 months, and the *VP instance* **presents** for the scheduled visit.
6. The *user* **asks** about swallowing again as well as about regurgitation and heartburn.
7. The *VP instance* **responds** that difficulty swallowing and regurgitation both occur several times a week; there is no heartburn.
8. The *user* **orders** a barium swallow diagnostic test
9. The *tutor agent* **blocks** the action, **responding** with the following message (in red) to the user:

   Precondition(s) for barium swallow

   ONE OF:

   - SUSPICION OF A MECHANIC OBSTRUCTION
   - SUSPICION OF A MOTILITY DISORDER

10. The *user* **posits** 'motility disorder' as the hypothesized disorder (using the diagnosis/hypothesis button for 'Esophageal motility disorder').
11. This action is **permitted** by the *tutor* because its preconditions have been satisfied (namely, the patient is known to have dysphagia, which is difficulty swallowing). Figure 1 shows the view of the user's interface at this point in the scenario.
12. The *user* again **orders** a barium swallow diagnostic test.
13. This time the *tutor* **permits** this action.
14. A *lab technician agent* **produces** numerical results for this test, **passes** them on to a *specialist agent*, who **returns** the results and an interpretation: "Subtle narrowing of LES. No dilated esophagus." The result and the interpretation are recorded in the patient chart.
15. The *user* decides that it is still too early to intervene, sends the patient home and

schedules another follow-up visit in 12 months.

16. When the *VP instance* **presents** at 32 months, difficulty swallowing and regurgitation have increased in frequency and severity, as the user can see by comparing the values for this visit to the values from the last visit (accessible by the dated tabs).
17. The *user* **orders** the diagnostic tests EGD, barium swallow and manometry, in that order.
18. The *tutor* **permits** all of these moves and, since maximum tutoring support is enabled, the preconditions for each of these actions are shown when the user carries out the action, reinforcing the teaching points related to prerequisites for good clinical practice.
19. The *lab technician* and the *specialist agents* **return** results of tests.
20. The *user* studies the results, and **posits** a diagnosis of achalasia.
21. The *tutor* **accepts** this diagnosis, once again, as in 18 above, displaying the preconditions for this action.
22. The *user* then **prescribes** BoTox as a treatment
23. The *tutor* **permits** this action (its precondition being a definitive diagnosis of achalasia).
24. The *user* decides he has had enough tutoring and **turns off** the tutor. The tutor's "unspoken" reaction to all the moves in the scenario can, however, still be seen by the user following the scenario in a log generated by the system.
25. The *user* **schedules** a visit in a month (month 33) for a follow-up.
26. At that time, the *VP instance* is **asymptomatic**, meaning that the BoTox was successful. However, BoTox's effects are only temporary, so there will be regression.
27. The next **follow-up** is in 9 months (month 42). The *user* **interviews** the patient who **responds** that she has significant difficulty swallowing.
28. The *user* **orders** the surgical procedure Heller myotomy, which cuts the lower esophageal sphincter to permit food to pass.
29. Three months later (month 45) is the next **follow-up**: the *user* asks about swallowing and also heartburn (since the hypotensive sphincter that results from Heller myotomy often leads to GERD). The *VP instance* **responds** that swallowing is fine but heartburn has developed.

In the continuation of the scenario, the user should treat the heartburn. Depending on this patient's inherent predispositions, her achalasia might regress, meaning that her lower esophageal sphincter would become tight again, reversing the GERD but reintroducing difficulty swallowing; alternatively, the achalasia might not return, meaning that the patient will have GERD for the rest of her life.

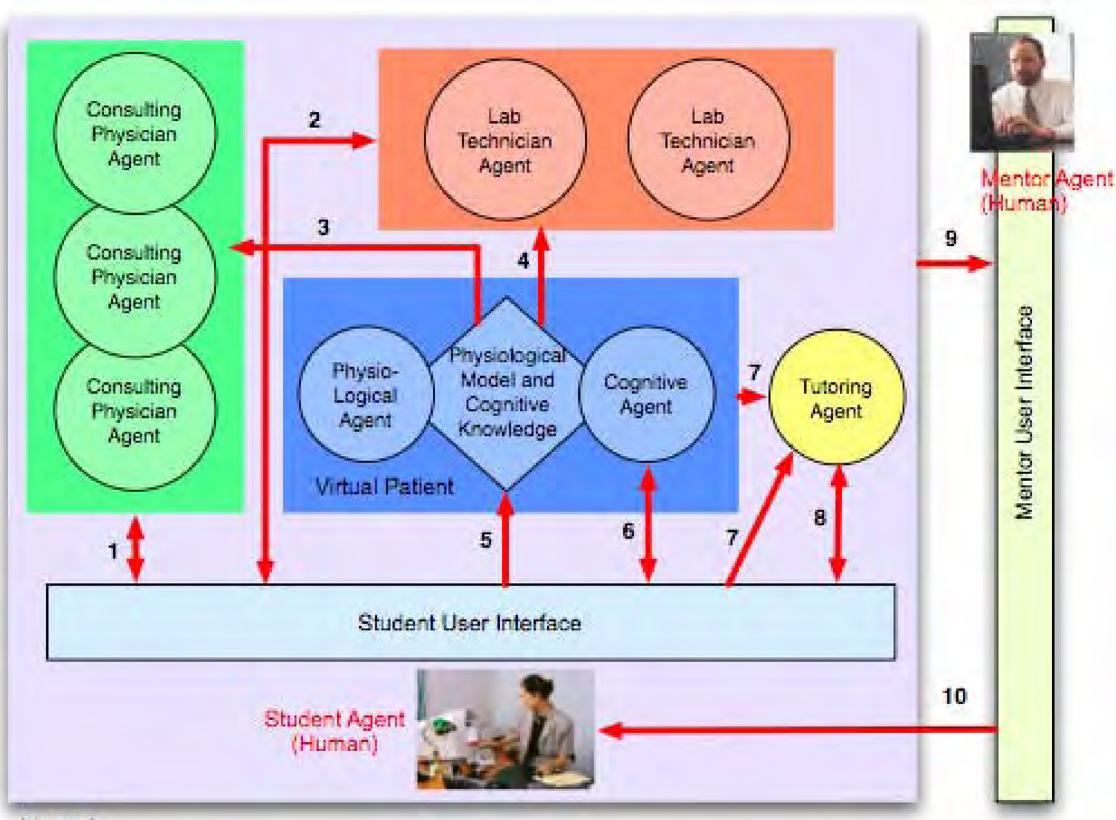
As this example illustrates, one of the novel aspects of the MVP system is how open-ended the scenarios are. An informal run-through with third-year medical students confirmed our assumption that the lack of a rigid structure – typically imposed by multiple-choice type learning environments – was both a surprise and a boon for students, since it so much more closely mimics what they will be required to do in clinical practice.

The Agent Network

A variety of views co-exist on what constitutes an intelligent agent. We divide our agents into high-level and low-level as follows.

High-level agents include actual humans, simulated humans (specialist consultants, lab technicians, the virtual tutor), and the physiological and cognitive sides of the virtual patient, which we choose to model as separate high-level agents. High-level agents are viewed as complex objects that can include models of the world and of self, including character traits, goals and plans. They also include various processing capabilities, such as decision making, language understanding and modifying a physiological profile. Figure 3 presents the high-level agents in the MVP network and the types of communication among them.

Low-level agents are, for us, not objects but processes. They may represent the specific capabilities of various high-level agents (e.g., language generation) or events in the world external to the high-level agents (e.g., exposure of a virtual patient to an environmental toxin). Low-level agents can be used to simulate individual physiological processes, both normal and pathological (i.e., diseases). Low-level agents realize a broad range of entities, from surgeries that affect the VP's physiology and anatomy, to external stressors that affect its cognitive and physical state, to individual physiological and pathological processes, to system functionalities that support simulation and tutoring goals (the patient creation agent, the information visualization agent, etc.).



Legend

1 From student: requests for information; from consultants: advice
2 From student: requests for labwork; from lab technicians: test results, augmented by interpretation by specialist
3 Consulting physicians get data from the physiological model of the virtual patient
4 Lab technicians get data from the physiological model of the virtual patient
5 Student can administer treatment, simulated by changing the physiological (and/or anatomical) model of the patient
6 From student: questions about symptoms etc.; from cognitive agent: responses in language
7 The tutoring agent collects information about the diagnostic and treatment processes
8 From the tutoring agent: advice, warnings, answers to user questions; from student: questions, responses to queries
9 The mentor observes the diagnostic and treatment as well as automatic tutoring processes
10 The mentor communicates directly with the student

Figure 3. The Maryland Virtual Patient (MVP) As Member of Its Multi-Agent Team that Includes Human and Simulated Agents

In addition to being categorized as high-level or low-level, the agents in the MVP network can also be characterized as deterministic or cognitive. In our approach, deterministic agents model physiological, environmental, chemical, etc., processes, while cognitive agents model sentient human behavior. Deterministic agents do not possess models of desires and intentions, they cannot be aware of other agents around them, cannot negotiate, cannot delegate and cannot have their decisions be influenced by purely agent-internal causes. For example, a disease agent can only advance the progression of the disease in accordance with the changing values of specific physiological properties of the virtual patient and the passage of time. Of the traditional triad of factors involved in modeling agency – beliefs, desires and intentions (see, e.g., Rao and Georgeff 1991) –

deterministic agents can be said to possess only the first, since it is plausible to interpret the knowledge of property values as the beliefs of an agent at a point in time.

The only MVP agent that is both high-level and deterministic is the physiological side of the VP. In our chosen application, there is no reason to model the physiological sides of the other simulated high-level agents in the system – the lab technicians, the consultants or the tutor. Note that there is a channel of communication between the physiological and the cognitive side of the virtual patient agent. It consists of the cognitive agent being able to trigger physiological processes and to receive from the physiological agent sensory information about symptoms (e.g., pain or difficulty swallowing). For example, swallowing is a deliberate conscious act, but only in its first stage: after the voluntary act of swallowing has been initiated (the gulp), the rest of the processes needed to get the bolus from the mouth to the stomach are deterministic physiological – though not pathological! – processes.

Two special agents in the MVP environment, the scheduler and the executor, support the operation and interaction of other agents.

In the subsections to follow we briefly describe the core agents in the MVP network, the data that supports their functioning, and their interactions with other agents in the network.

The virtual patient is the most complex agent in the system. This is because we model both its body and its mind. In other words, the virtual patient is for us a “double” physiological and cognitive agent. The physiological agent is a simulation of physiological and pathological processes. The cognitive agent combines the capabilities of perception (specifically, proprioception and language understanding), goal- and plan-based reasoning and action (decision-making and language generation).

The operation of the physiological agent is not directly controlled, though it can be influenced, by decisions and choices of the cognitive agent. Examples of such influences are lifestyle preferences, like smoking, regular exercise, and diet. The operation of the cognitive agent can, in turn, be influenced by the physiological agent. Indeed, the cognitive agent’s choice of goals, and the choice of plans for their attainment, will be influenced by the physiological agent’s physical state (e.g., disease, fatigue) and mental state (e.g., stress).

The physiological agent models the body as a collection of anatomical objects and physiological processes, including both normal and pathological ones (diseases). Whenever possible, disease processes are modeled as causal chains of component events (and, implicitly, states). These causal chains are encoded as complex events – i.e., scripts – in the system’s static knowledge (specifically, in its underlying ontology). However, in many cases, medicine does not at this time possess sufficient knowledge about biochemical mechanisms of disease progression to allow for the construction of completely causal scripts. This means that disease scripts must often contain a combination causal chains and empirical knowledge about the progression of a disease -- what we refer to as clinically derived “bridges”. In the current implementation, when it is not possible to encode a causal chain, the progression of the disease is divided into clinically relevant conceptual stages, and a set of value ranges of relevant physiological

properties is encoded for the beginning and end of each stage. During simulation, values for interim time points are established through interpolation.

In our knowledge acquisition work we have found that expert clinicians like to express their opinions in terms of probabilities, e.g., "X% of patients develop stage N of disease D within M months of inception." Though the use of value ranges instead of single values reflects this state of affairs, in our current application system the probabilities do not actually play a central role. This is because the system is supposed to train future physicians, and this is best done when the instructor can create an inventory of virtual patients carrying a particular disease such that the patients display the full spectrum of disease manifestations, symptom profiles and responses to treatments, not only the most common ones. Presenting the student with a carefully crafted set of such patient instances permits all of the necessary learning points to be targeted without the need to spend years of real-world training waiting for the less common patients to present. A case in point: in the evaluation of the SHERLOCK II system, which teaches electronics troubleshooting, it was reported that technicians learned more from using this system for 24 hours than from 4 years of work in the field (Evens and Michael 2006).

Once the progression of a disease reaches the symptomatic (i.e., clinical) stage, the simulation reaches the cognitive side of the double agent. Through proprioceptive perception, the cognitive agent becomes aware of symptoms such as pain or difficulty swallowing. Note that the experiencing of symptoms varies widely across patients and, accordingly, cannot be directly linked to given physiological states. However, a fixed inventory of symptoms is associated with each disease and expected ranges of values for each symptom can be asserted for each stage of the disease.

While the physiological agent has been built as a result of the formal encoding of expert knowledge about the progression of diseases and response to treatments, the knowledge encoded for the use of the tutor agent concentrates on formalizing the what expert physicians understand to be best clinical practices – for example, what preconditions should occur for particular physician actions to be warranted and what the best ways of scheduling these actions are. Our work on the tutoring agent to-date has concentrated on these contentful issues much more than on the specific pedagogical choices (so-called "tutoring moves") that are at the center of interest in many intelligent tutoring projects in medicine and other areas (e.g., Evens and Michael 2006). At present, we implement only a small subset of choices available to a tutor. For example, different tutor settings allow more or less frequent interventions, the presence or absence of explanations for why certain suggested user actions were blocked, and so on. While we will continue to enhance the repertoire of choices in tutoring, our main goals are the adequate simulation of human physiology, the complete encoding of relevant best clinical practices, and the support of realistic natural language dialog between the user and the virtual patient.

The Physiological Agent of the VP

The VP physiology agent is modeled as a set of interconnected ontological objects representing human anatomy. Each object is described by a set of ontological properties and their associated value sets. Crucial among the properties are those that link the objects to typical events in which they participate. These events are usually complex – that is, include other, possibly also complex, events as their components. We call these complex events scripts, an example of which is swallowing. The swallow script involves

numerous anatomical objects in the VP carrying out various roles. For example, the esophageal segments that comprise the esophagus are composed of various types of tissue, muscle and nerves; the segments act as *instruments* of the peristalsis that forces the swallowed bolus toward the stomach.

All events within the simulation environment are interpreted as low-level agents. Swallowing, as such an agent, knows what role each property of each component part of the esophagus plays in the process of swallowing. Specifically, it knows what property values are normal and what effects to post after a normal swallow, and it knows what property values are abnormal and what effects to post after an abnormal swallow, depending on which abnormality was encountered. For example, a tumor in an esophageal segment decreases the size of the esophageal lumen and thus causes the symptom "difficulty swallowing" in the patient; the larger the tumor, the greater the difficulty swallowing, until the point where swallowing is blocked altogether. For the task of modeling esophageal diseases, having a normal, working model of the process of swallowing is useful, since asking a patient to swallow, and asking whether that triggered any symptoms, is a realistic component of a simulated office visit.
At first blush, it might seem preferable to have a maximally complete model of normal human anatomy and physiology before progressing to disease modeling, but we have found this not to be the case for three reasons:

1. Creating formal models of everything known about human physiology would require an unsupportable amount of time and resources.
2. Even if such models could be created, they would represent a grain size of description not needed for our applications.
3. Many of the processes of human physiology – both normal and pathological – are not understood by the medical community, meaning that modeling must anyway combine aspects of known causal chains and clinical observations that we call "bridges."

In short, all modeling in the MVP system is task-oriented, with both normal and pathological processes being modeled on an "as needed" basis. Achieving a useful balance between causal chains, bridges, and grain size could be considered the art of application-oriented modeling.

At any given time, the model of the normal human contains whatever normal anatomical and physiological knowledge was compiled to cover the diseases currently available in the system. So, although at present our virtual humans do not have a highly developed model of the circulatory system, as soon as we have completed the circulatory model – which is currently under development to support the modeling of heart disease – all virtual humans will be endowed with all the associated functionalities and property values.

Disease Agents

In the MVP system, diseases are modeled as processes (low-level agents) that cause changes in key property values of a patient over time. For each disease, a set number of

conceptual stages is established and typical values or ranges of values for each property are associated with each stage. Relevant property values at the start or end of each stage are recorded explicitly, while values for times between stage boundaries are interpolated; the interpolation currently uses a linear function, though other functions could as easily be employed.

A disease model includes a combination of fixed and variable features. For example, although the number of stages for a given disease is fixed, the duration of each stage is variable. Similarly, although the values for some physiological properties undergo fixed changes across patients, the values for other physiological properties are variable across patients, within a specified range. The combination of fixed and variable features represents, we believe, the golden mean for disease modeling. On the one hand, each disease model is sufficiently constrained so that patients suffering from the disease must show appropriate physiological manifestations of it. On the other hand, each disease model is sufficiently flexible to permit individual patients to differ in clinically relevant ways, as selected by patient authors.
For illustration, we take a simple disease model, the one for scleroderma esophagus. It is simple because the causal chains driving the disease are not known to medicine, making it necessary to model the disease purely in terms of clinical observations over time.[6] Table 1 shows the physiological properties that change over time. The values in each cell represent the values at the beginning of each stage: so a given patient might have basal lower esophageal sphincter (LES) pressure of 18 mmHg at the start of t0, 11 mmHg at the start of t1, etc. Legal value ranges are indicated, with defaults in parentheses. Properties for which no choice of range is provided (e.g., level of anti-nuclear antibodies) are fixed across all patients (subject matter experts saw no benefit to making such properties variable across patients). The symptom profile for patients with scleroderma esophagus is shown in Table 2, with value ranges and defaults represented using the same conventions.

Table 1. Physiological properties that change due to scleroderma esophagus

	t0	**t1**	**t2**	**t3**	**t4**
Level of anti-nuclear antibodies	0	1	2	3	4
Peristalsis efficacy	normal peristalsis	intermittent-peristalsis	aperistalsis	aperistalsis	aperistalsis
Basal LES pressure (in mmHg)	0-25 (15)	0-15 (10)	0-10 (5)	0-10 (3)	0
Erythrocyte sedimentation rate	0-15 (12)	5-20 (10)	20-30 (25)	31-40 (35)	41-50 (45)
Stage duration (in years)	1-5 (1)	1-5 (1)	1-5 (1)	1-5 (1)	1-5 (1)

Table 2. Symptom profile for patients with scleroderma esophagus

	t0	**t1**	**t2**	**t3**	**t4**
Raynaud's symptoms	no	yes	yes	yes	yes

[6] We choose a simple disease model here in order not to stray from the main thrust of the article: the agent network in the system. For descriptions of more complex disease models, see McShane *forthcoming* (for achalasia) and McShane *submitted* (for GERD).

Skin tightness	0	0-.3 (.1)	.1-.4 (.3)	.3-.6 (.4)	.5-.8 (.7)

Scleroderma esophagus is a disease whose path cannot be altered by interventions by any outside agents. However, the physiological changes it causes give rise to another disease: GERD. Specifically, when the basal LES pressure drops below 10 mmHg (typically in the t2 or t3 stage of scleroderma esophagus), the amount of stomach reflux permitted by the hypotensive sphincter will be sufficient for irritating processes of the esophageal lining to begin – i.e., GERD. There is nothing in the disease model for scleroderma esophagus about GERD because there need not be: the initiation of GERD is handled by an object-oriented condition on the LES that is triggered when the LES pressure drops below 10 mmHg.

In the simple scleroderma esophagus model, there are few parameterizable values; however this is not so for all diseases. For example, GERD has six radically different clinical manifestations – from harmless but symptom-inducing irritation of the esophageal lining to esophageal cancer. Which path a patient takes depends upon a number of parameters, including genetic predispositions, lifestyle habits, and certain physiological features, like LES pressure, that can be affected by outside agents.

The Cognitive Agent of the VP

As mentioned above, some diseases, like scleroderma esophagus, cannot be affected by an external agents, but many others can. The cognitive agent can affect disease progression by choosing **lifestyle habits** and by **compliance or non-compliance with medication regimens.** Let us consider the example of GERD, one of whose causes is a hypotensive LES.[7] The LES can be made more hypotensive by lifestyle habits like consuming caffeine and eating chocolate. The more hypotensive the LES, the higher the daily acid exposure of the esophageal lining and the faster the disease progression, based on the causal chain of events used to model GERD (see McShane et al. submitted).

Consider a patient whom we'll call Patient A. This patient has a basal LES pressure of 4 mmHg, he engages in the GERD-irritating habit of consuming caffeine, and he is genetically predisposed to a form of GERD that progresses to the level of producing esophageal erosions but never produces ulcers, cancer or other complications. If the patient does nothing to improve his disease course, it will proceed as in Figure 4, which shows just four of the dozen or so property values relevant for GERD.[8] However, if Patient B stops his caffeine habit in month 5 of the disease, when his heartburn is already at a .5 on the scale of 0-1, his LES pressure will increase slightly, decreasing acid exposure and producing a disease course as in Figure 5, where heartburn severity reaches

[7] The other cause is transient relaxations of the LES, which, like a hypotensive LES, permit excessive exposure of the esophageal lining to acidic stomach contents.

[8] To briefly explain the properties shown: when preclinical esophageal irritation reaches its maximum, the clinical (i.e., symptom-producing) stage of the disease begins, in this case, causing heartburn; when the mucosa depth reaches 0, erosive esophagitis begins (in the simulation an erosion object would be instantiated); and when the mucosa depth reaches 0, the erosion has reached its maximum depth – any further eroding would constitute an ulcer, which this patient is not predisposed to get. In this and subsequent figures, when different property values have different scales of measurement, we normalize them to make their changes over time most visible in the graph. We call the resulting scale an abstract scale.

its maximum at around month 10 rather than month 9. Clearly, for this patient, lifestyle modification is not enough to stop or reverse the disease course, but it causes some modification of disease path.

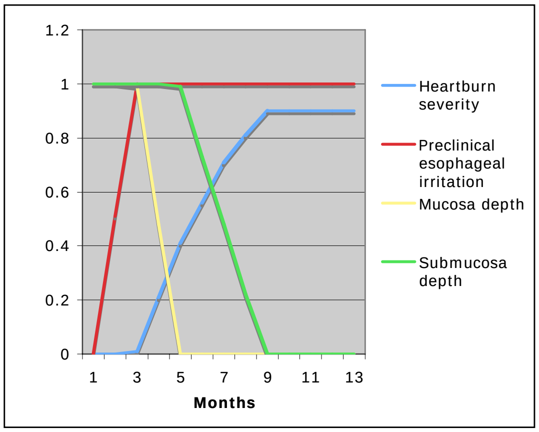


Figure 4. Patient A with no interventions.

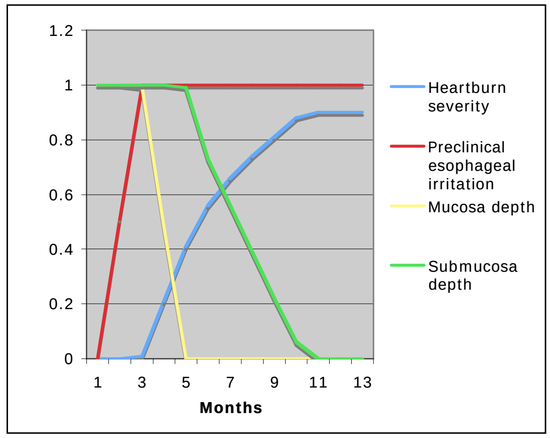


Figure 5. Patient A with lifestyle modifications in month 5.

Now assume that in month 5 the patient decides to see a doctor and is put on medication that, for him, is effective. (See Figure 6.) Assume further that he takes that medication for three months then stops taking it. During the period when he is taking it, his disease heals and remains healed, but as soon as he stops taking it, the disease begins to progress again, first asymptomatically (in the preclinical phase), then symptomatically.

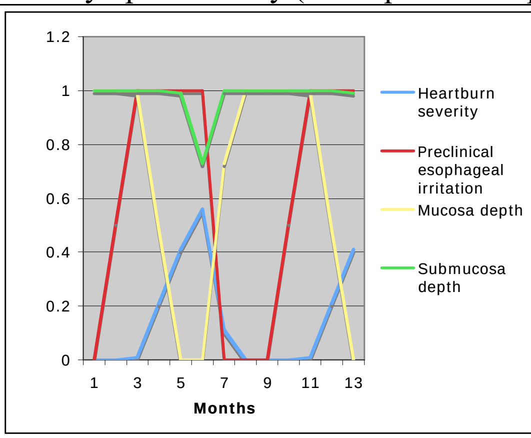


Figure 6. Patient A with medication starting in month 5, ending in month 8.

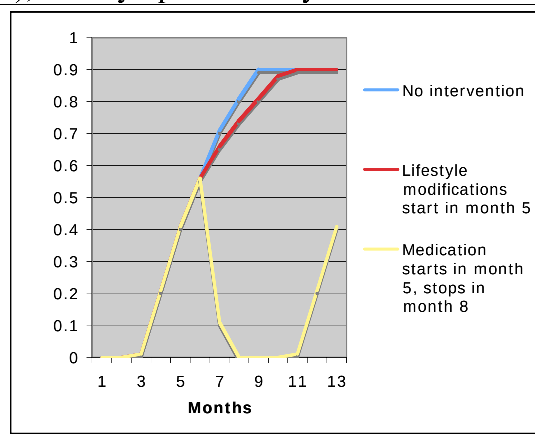


Figure 7. Patient A: heartburn severity under different treatment scenarios.

Figure 7 compares the symptom-related experience of the cognitive agent of Patient A under these three scenarios using just the symptom "heartburn severity".

The behavior of the cognitive agent with respect to these choices is determined by an inventory of personality traits that can be selected by the patient author. Our initial inventory, which has been sufficient for esophageal diseases, includes pain threshold, willingness to seek medical help and compliance with medication regimens. However, the group of diseases we are currently in the process of modeling – cardiovascular diseases – relies much more heavily on personality and lifestyle factors, which will

require an increase in the inventory of ways in which the cognitive agent can influence the operation of the physiological agent.

The above discussion relates to only those components of the VP cognitive agent that model influences on the VP physiological agent. Of course, the cognitive agent is also responsible for a wide variety of sensory and proprioceptive perception, reasoning, decision making and inter-agent communication processes. Figure 9 illustrates the component processors and static knowledge resources of the VP cognitive agent. The figure includes both stable components that have largely been developed and components under development at the time of writing.

Figure 9. The Processors and the Static Knowledge Modules of the Cognitive Agent component of the virtual patient. Work is ongoing on all these modules. The constraints of the application help to keep the inventory of goal and plan types, the character traits and the attention resources relatively small. The general world knowledge (the agent's ontology) and its lexicon are close to being complete. The discourse and dialog processing is a core direction of work in 2007-8.

The User

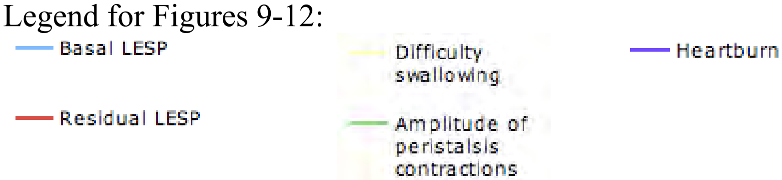
At the beginning of a simulation session, the system presents the user with a virtual patient about whose diagnosis he initially has no knowledge. The user then attempts to manage the patient by conducting office interviews, ordering diagnostic tests and prescribing treatments. The basic strategies of patient management were illustrated in the example in Section 2.2.

Answers to user questions and results of tests are stored in the user's copy of the patient profile, represented as a patient chart (see left pane of Figure 1). At the beginning of the session, the chart is empty and the user's cognitive model of the patient is generic – it is just a model of the generalized human. The process of diagnosis results in a gradual modification of the user's copy of the patient's profile so that in the case of successful diagnosis, it closely resembles the actual physiological model of the patient, at least, with respect to the properties relevant to the patient's complaint. A good analog to this process of gradual uncovering of the user profile is the game of Battleship, where the players gradually determine the positions of their opponent's ships on a grid.

At any point during the management of the patient, the user may prescribe treatments.[9] In other words, the system allows the user not only to issue queries but also to intervene in the simulation, changing property values within the patient. Any single change can induce other changes – that is, the operation of an agent can at any time activate the operation of another agent.

Consider the example of Patient B, who suffers from achalasia, a disease that increases the basal and residual pressure of the LES, making swallowing increasingly difficult. The four graphs below show different outcomes based on different interventions by the user.

Legend for Figures 9-12:



[9] In this discussion, we disregard the tutor agent that can, in some pedagogical configurations, prevent a user action.

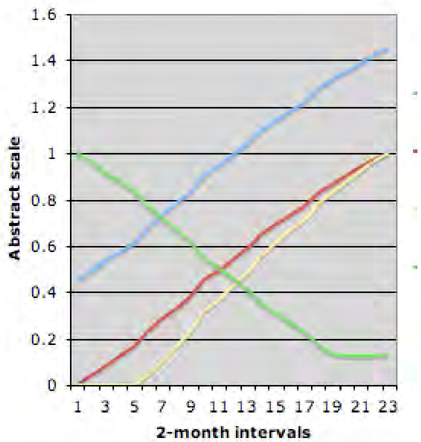


Figure 9. Patient B with no interventions.

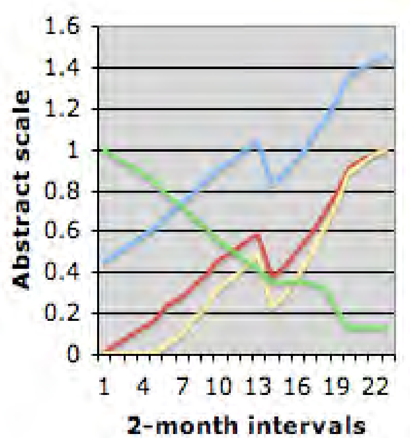


Figure 10. Patient B with BoTox in month 26.

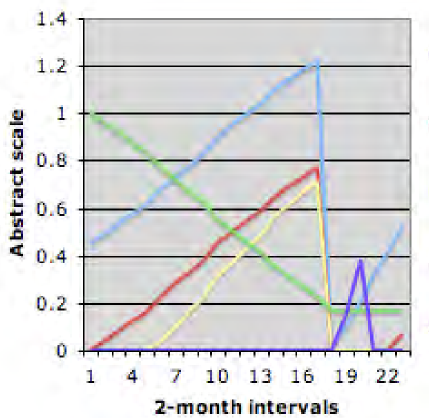


Figure 11. Patient B with Heller myotomy in month 34.

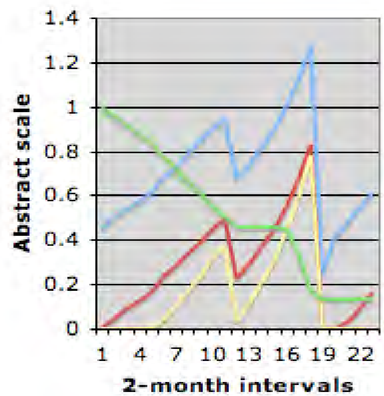


Figure 12. Patient B with BoTox in month 22 and pneumatic dilation in month 36.

Figure 9 shows the disease progression with no intervention. Figure 10 shows what happens when BoTox is administered: this treatment, if effective for the given patient, lowers basal LES pressure temporarily, but as it wears off the disease path returns to where it would have been if the treatment had not been given. Figure 11 shows intervention with the surgical procedure Heller myotomy in month 34. This procedure cuts the LES, typically leaving it with a pressure of near 0. With such a low LES pressure, a patient generally experiences excessive reflux and begins to suffer from GERD, as shown by the spike in heartburn around month 38. The reason the heartburn does not continue indefinitely is that the author of this patient instance decided that Heller myotomy would not provide definitive treatment for this particular patient; as a result, the tightening of the LES that is characteristic of achalasia again takes hold over time. By month 44 the patient's LES pressure has increased sufficiently to stop the GERD symptoms and, if we were to track the disease path even longer, we would see that his difficulty swallowing would continue to increase. Figure 12 shows two interventions: first, BoTox is administered and provides a temporary decrease in symptoms; then, when symptoms return to a high level, the endoscopic procedure pneumatic dilation is carried out, which tears the LES using an inflated balloon. This procedure, like the Heller myotomy, can be definitive or not definitive, depending on the patient's predispositions.

The four scenarios above illustrate but a handful of the thousands of scenarios one could create using Patient B, since any of the treatments could be administered in any combination at any time. In addition, there are *incorrect* treatments that could be administered which might be harmless or might worsen the patient's condition, creating a more complicated case that the user must manage. Moreover, in addition to Patient B, who has a specific inventory of disease-related property values, hundreds of other patients with this disease could be authored, making the scope of cases truly wide and differentiated.

Virtual Medical Personnel

Currently, the agents simulating medical personnel in the system include lab technicians, specialists that interpret test results and specialists that carry out procedures. In the future, we will include virtual user assistants to carry out physical exams.

Virtual lab technicians are agents that know which property values each test targets and check those values in the current state of the virtual patient, returning the values as strings. Test-interpretation specialists are endowed with a data set that associates given property values and combinations of property values with given interpretations (strings).

Specialists that carry out procedures are currently modeled in a rather simplified way: the results of the various tests and procedures, if carried out on the patient at various times, are pre-recorded and those results, in a sense, encapsulate the work of the specialist. In future, we will endow the specialist agents with a sufficient number of features to permit the autonomous computation of the quality of their results.

Environmental Agents

Environmental agents are a diverse set of low-level agents external to the VP that can have a profound effect on virtual patients. We expect many of these to be launched randomly within parameters set by the patient author. Two examples of environmental agents are stressors and the exposure of the VP to entities, like viruses and toxins that can cause disease processes. For example, for patients with heart disease, both long-term and acute stressors significantly affect disease course. To model the influence of environmental agents like stressors, the MVP environment will initiate, at points in time prescribed by the VP instance author, the operation of agents that raise the VP's stress levels (e.g., a car accident, divorce). Although participants in some such events could, in principle, be modeled as high-level agents (e.g., the driver of the car that caused that accident or the microorganism whose activities caused a disease), there is no practical reason for doing so in our applications. Instead, all environmental agents are interpreted as low-level agents.

The Tutor Agent (The Virtual Mentor)

The tutor agent, the virtual mentor, is one of two pedagogical agents in the environment, the other being a human mentor whose potential participation we will not detail here.

Much of the knowledge of the tutor agent derives from what we call "preconditions for good clinical practice". These are formally specified conditions that should be satisfied before a given hypothesis or diagnosis is posited, or before a given test or treatment is launched. For example, before carrying out a Heller myotomy, a definitive diagnosis of achalasia must be made; and in order for a definitive diagnosis of achalasia to be made, the following preconditions must hold:

ALL OF:
1. ONE OF:
    - bird's beak ; one interpretation generated from barium swallow test
    - LESP > 45 ; basal LES pressure
2. aperistalsis
3. ONE OF:
    - dysphagia ; difficulty swallowing
    - regurgitation

So, if the tutor agent is enabled and the user attempts to perform a Heller myotomy without fulfilling the necessary preconditions, the tutor will generate a message. The nature of the message depends upon the tutoring option selected: at the highly informative end of the scale, the tutor can provide all fulfilled and unfulfilled preconditions, with the former being in green and the latter being in red; at the most uninformative end of the scale, the tutor can respond "Invalid action", leaving the user to figure out what needs to be done to fill the lacunae.

Certain tutor settings make it impossible for the user to carry out "bad practice" actions. In this way, the tutor not only affects the human user, it also affects the virtual patient run by making impossible certain paths of disease progression and healing. Consider again the achalasia patient we called Patient B. With the tutoring disabled, the user is permitted to treat the patient with BoTox in month 18, but with tutoring enabled, the first time the user could administer BoTox would be in month 26, which is the earliest time that all the preconditions for administering it are fulfilled.

Another way in which the tutor will, in the future, be able to affect both the user and the virtual patient is by suggesting what to do next when the user asks for help. This functionality requires that the tutor be endowed with more and different knowledge than that needed for the current preconditions-driven tutoring method. Specifically, it will need to decide – based on its knowledge of the good practices for diagnosis and treatment – which of many moves available at any given time in the process is optimal and be able to convey the reasoning behind that decision to the user.

Much of the work on intelligent tutoring systems concentrates on the implementation of a variety of pedagogical "moves" (Evens and Michael 2006). We have already implemented a subset of such moves by introducing different levels of tutor intervention, and will integrate more as work proceeds.

An automatic tutor can perform a variety of useful tasks. For example, it can track the decisions the human user takes and suggest alternative (better) courses of action, alert the user to errors or unfulfilled preconditions for various actions, and so on. One of the most important roles of a tutor in a clinical setting is providing the human user with advice about *what to do next*. In order for the virtual tutor to decide what should be done next, it must combine context-independent knowledge with context-specific reasoning.

The context-independent knowledge of our virtual tutor covers, non-exhaustively:

1. **high-level best clinical practices**: e.g., that the physician should interview the patient before all else (in non-critical situations); that he should posit a working hypothesis or diagnosis before ordering tests and procedures; that each test and procedure should be ordered only if the patient meets certain objective criteria warranting it, etc.;
2. **diseases**: their signs and symptoms over time; the chief complaints they give rise to; other diseases with overlapping signs and symptoms; variations in disease manifestation across patients; what constitutes sufficient evidence to hypothesize, clinically diagnose or definitively diagnose each disease;
3. **diagnostic tests**: the specific criteria that should be met before ordering them; what they test for and what kinds of results they return; potential complications and their frequency; the availability and time frame of each test;
4. **interventions**: the specific criteria that should be met before ordering them; their projected outcome; potential side-effects and the frequency of those side-effects; the availability of qualified specialists to carry out various procedures.

Context-specific reasoning combines this static knowledge with knowledge about the patient and the student that the tutor compiles during the course of the given simulation run. Such dynamically created knowledge includes: (a) the current state and past history of the patient, as elicited by the student through patient interviews, (b) each of the past actions of the student (questions asked, tests ordered, interventions performed, hypotheses and diagnoses posited), and (c) the content of any past interactions between the tutor and the student (the student asking questions and the tutor answering; the tutor intervening to stop the student from making an ill-advised move, etc.). Note that the tutor does not have omniscient knowledge of the patient's physiology since no physician can work from that unrealistic starting point – it knows about the patient exactly what the student knows about the patient. The difference between the tutor and the student in this respect is that the tutor embodies the knowledge of best clinical procedures, as possessed by expert clinicians and encoded by knowledge engineers in the knowledge resources of the system.

The tutor can use its combined static knowledge and reasoning capabilities to function in a number of tutoring modalities, including:

1. providing step-by-step commentary about whether each move is clinically appropriate and why;
2. stopping the student before he carries out clinically inappropriate moves;

3. showing the student how to fulfill the unfulfilled preconditions for a given move;
4. suggesting which move(s) should be carried out next and why.

The first three of these tutoring modalities – as well as a number of variations on the theme – are available in the first release of MVP, and the last one is currently under development. Let us consider one scenario that highlights the adaptive nature of the tutor's reasoning capabilities:

1. The patient presents with the chief complaint "occasional difficulty swallowing."
2. The student interviews the patient, finding out that the only other symptom is occasional mild chest pain.
3. The student hypothesizes the disease 'achalasia.'
4. The student orders an esophagogastroduodenoscopy (EGD). Negative results for the EGD are returned by the lab technician and specialist agents.
5. The student then orders a barium swallow. It reveals a slight narrowing at junction of the stomach and the esophagus (the GE junction).
6. At this point the student asks what he should do next.

During the simulation, whether or not the step-by-step commentary function is enabled, the tutor evaluates each move the student makes, saving the evaluations in a log that can be reviewed later by the student and/or teacher. The tutor would approve of step 2 as long as sufficient questions about related diseases have been asked. For example, since the patient complains of chest pain, questions about other symptoms of heart disease should be asked, since chest pain is a primary symptom of heart disease. The inventory of symptoms that should be asked about is *dynamically generated* based on patient responses to each symptom question: the more symptoms the patient has, the more associated symptoms must be asked about, to rule out similarly presenting diseases.

The tutor would approve of step 3 but would suggest that a better hypothesis would be "motility disorder," which is a superclass of achalasia, since there is no evidence at this point in the proceedings suggesting which specific motility disorder this actually is. The tutor evaluates whether a hypothesis is reasonable on the basis of the filler of the ontological property SUFFICIENT-EVIDENCE-TO-HYPOTHESIZE, which is defined for each disease and class of diseases. If the above property has the value DIFFICULTY-SWALLOWING listed, this is sufficient grounds to hypothesize a MOTILITY-DISORDER, and is also sufficient grounds to hypothesize its child, ACHALASIA, but hypothesizing the more general disorder is better clinical practice in the absence of further evidence – a generalization known by the tutor.

The tutor would approve of step 4 on the basis of the clinical practice of *first ruling out any alarm signals* (i.e., potential immediate causes of danger). Difficulty swallowing can suggest a tumor, which could be cancerous, so the first action should be to rule that out. The need to rule out cancer in the presence of the symptom DIFFICULTY-SWALLOWING is recorded using the property TRIGGER-ALERT in the ontological description of DIFFICULTY-SWALLOWING, as follows:

DIFFICULTY-SWALLOWING
  TRIGGER-ALERT   TUMOR (LOCATION ESOPHAGUS)

Fillers of the property TRIGGER-ALERT should be pursued first, even if the related condition is unlikely and does not represent the current working hypothesis. Ontologically recorded knowledge about EGD includes the fact that it can detect tumors. It also includes the fact that DIFFICULTY-SWALLOWING, by itself, is a sufficient condition to order an EGD. As a result, the tutor determines that the student has acted correctly in ordering an EGD as the first study.

Whereas the EGD was ordered in step 4 in order to rule out a potential problem (tumor), the test ordered at step 5 is intended to *provide evidence to confirm the current hypothesis.* Ordering tests to confirm a hypothesis represents clinically correct behavior, since for many diseases (including achalasia) a diagnosis must be made before any treatment can be launched.

It is not surprising that the student would ask for help after the results of the barium swallow are received. He or she had a hypothesis but that hypothesis was not confirmed by the test ordered: the barium swallow would have to have shown a finding known as "bird's beak" at the GE junction – rather than just a slight narrowing – if achalasia were to be diagnosed. The student has not fulfilled the preconditions to order any more tests; and since a diagnosis of achalasia has not been confirmed, no treatments can be launched. *What should I do now?*, the student asks the tutor.

At this point, the tutor reviews the path taken so far and searches for any points at which other moves might have been taken. Formally, this means comparing the preconditions for all the available actions with the knowledge available at the time of each move. In this case, the student acted similarly to the way the tutor would have acted except for the following: (a) as described above, the tutor would have posited a motility disorder, not achalasia; and (b) the tutor might have sent the patient home after receiving the results of the EGD, since the patient was in no danger and its symptoms were very mild (whether to go ahead with the barium swallow or wait and see how the symptoms progress is a judgment call; an experienced diagnostician might have guessed that there would be insufficient evidence to diagnose any disease at this point, based on the patient interview). The tutor would point out to the student the slight deviations from how it, the tutor, would have handled the case, and would tell the student that at this time the best thing to do is to send the patient home, with a recall in a few months or if symptoms become significantly more pronounced.

This option of "wait and see" is one aspect of the system that both makes it extremely open-ended and trains students to do something that they are known to find uncomfortable – namely, not take immediate action. Most training environments for medical students tell students that they must do something immediately and give them a choice of what to do. However, physicians frequently must simply wait to see how a disease plays out, reassuring the patient and sending him home. In the scenario above, this is exactly the advice the tutor would give the student. The point, however, is that all

of the tutor's decisions are based on comparing its static knowledge with the dynamically changing evidence available about the state of the patient and the knowledge of the student.

Architecture and Control Agents

This section presents an informal, content-oriented description of the operation of the MVP. The MVP system uses agenda-style control that operates on data stored in the fact repositories of each of the high-level agents.[10] Each high-level agent – i.e., virtual human – has its own blackboard, which represents its current inventory of property values. The agenda is implemented as a temporally ordered list of agenda slots. Each agenda slot is a set of event instances scheduled for execution at a particular time.[11] The frequency of agenda slots (the clock cycle period) can be manipulated in the MVP system: the temporal scale of the simulation can be changed in a broad range between milliseconds and years.

Simulated events (scripts) can be simple or complex. A complex event contains subevents whereas a simple, atomic event does not. Whether an event is modeled as simple or complex generally depends upon the selected grain-size of description: for example, in a non-medical application the event of swallowing might be modeled as a simple event, but in our medical application it includes dozens of subevents describing component physiological processes that can be affected in various ways by various disease processes.

From the standpoint of MVP operation, the most important properties of event instances are:

1. **Preconditions-effects**, whose values are effects of the event, each coupled with a set of its own preconditions. For example, for event E, if precondition A holds, then the event will be sent to the scheduler to be rescheduled in the next agenda slot (this is the way we model continuing events), whereas if precondition B holds, changes to values of specified properties will be posted in appropriate fact repositories. Preconditions typically involve one or more property values of the agent, of the world or the simulation process itself; the completion of another event is another typical type of precondition. Effects typically involve changing property values of the VP or scheduling new events.
2. **Subevents** are defined for complex events. Subevents can themselves be complex events, with no conceptual limit on the number of nesting levels. The subevents of an event are organized in a directed graph, whose nodes represent the subevents and whose edges represent a temporal ordering of the subevents. Whenever the system

[10] Each agent has its own fact repository that reflects its private knowledge (beliefs) about the world and includes a record of current and past events involving the agent or known to it. Thus, for example, the user's fact repository includes the subset of the VP physiology profile revealed to the user and the record of past events recorded in the patient chart. The complete VP physiology profile is the main fact repository in the system – all the changes to the VP properties are recorded there.

[11] For purposes of simulating parallelism, agenda slots have been implemented as last-in-first-out queues.

knowledge allows, the temporal precedence relations are also marked as causal connections, which subsume the temporal ordering.

3. **Duration**. For simple, atomic events, the stated duration is an important (though not the only) factor that determines the time at which its effects are posted by the executor. The duration of complex event instances is computed as the sum of the durations of all the subevents on a particular path in the subevent graph.

The control agents driving the MVP operation are the scheduler, the executor and a (potentially large) set of "demon" agents, which are procedural attachments to properties in the high-level agents that trigger events as a result of specified changes in the values of the properties to which they are attached. For example, if the pressure of a virtual patient's lower esophageal sphincter drops below 10 mmHg (as by a surgical intervention), this state of affairs is sensed by the corresponding demon, which triggers the action of scheduling the complex event "GERD" (and as a result of this, the VP physiological agent gets this disease). The MVP system, thus, is a mixed expectation- and data-driven system. The demons trigger reactive, data-driven events. Expectation-driven events (scripts and plans) are encoded in the ontology. The scheduler and the executor manipulate both kinds of events.

The **scheduler** performs the following operations:

1. Accepting requests for events to be put on the agenda (see below for the inventory of types of requests);
2. Placing those events in appropriate time slots on the agenda; and
3. Removing events from the agenda; removal of an event can be caused by (a) the executor having determined that the preconditions of none of its effects are fulfilled; or (b) a direct request from an agent (e.g., the patient cancels a follow-up visit).

Any rescheduling operation – e.g., moving an appointment to a different date – is interpreted as a removal action followed by a scheduling action.

The **executor** performs the following operations for each event it processes:

1. Checking the preconditions of each of the effects of an event, and
2. Posting each of the effects whose preconditions hold.

One of the possible effects of an event is sending a request to the scheduler to (re)schedule it. The precondition associated with this effect involves the stated duration of the event and, optionally, property values of the patient or the world at the time when the precondition is evaluated.

The latter constraints are included to alleviate the frame problem – to account for potential changes to the world caused by other agents (events). Theoretically, the rescheduling should occur for the next agenda slot, and the rescheduling process should repeat until the agenda reaches the slot corresponding to the end of the duration of the

event (unless external influences expressed in additional preconditions dictate otherwise). Considering that the clock cycle period in the MVP can be on the order of milliseconds, this will lead to inefficiencies. Therefore, in practice, we assign a default clock cycle period to each kind of event (e.g., one second for esophageal peristalsis; one month for a stage in the GERD event, etc.) and specify it in the rescheduling request. This, in turn, leads to constraints on the possible times of occurrence of external events: for example, the author's suggestion that a stressor be scheduled for a given patient at time T might be "relaxed" by the scheduler, such that it is scheduled at T +/- some amount, such that it does not occur between planned agenda slots. As yet, such efficiency-oriented constraints have not been shown to affect the realism of the simulation or to limit the pedagogical impact of the system.

Requests for putting events on the agenda can be sent to the scheduler in several ways:

1. The scheduler can have a "standing request" for scheduling normal physiological events (breathing, heartbeat, etc.) at appropriate intervals. One of the effects for such events will be to schedule another instance of this event once the previous instance finishes executing.
2. The VP instance author may stipulate that certain events should be scheduled at specific times. For example, the author requests the scheduling of disease scripts (agents) and environmental agents (e.g., exposure to a toxin) for appropriate times.
3. User interventions – the asking of questions, the ordering of tests or procedures, the scheduling of a VP's follow-up visit – are requests to the scheduler that those events be put on the agenda.
4. Actions by the VP cognitive agent, like ceasing to take some medication, are sent to the scheduler.
5. Effects of actions being executed by the executor may include sending a scheduling request. For example, when the event of the user asking the VP agent a question is executed, the effect is scheduling the events of the VP agent understanding this question and formulating the answer to it. Another example is scheduling the next subevent in a complex event once the previous event has completed its execution.
6. Requests for scheduling events can appear as a result of the operation of demons.

Medical Knowledge Creation and Formalization

The core medical knowledge in the system covers human physiology, pathology, relevant farmacology and a record of physician expertise – the "best clinical practices." This knowledge is needed for the support of simulating disease progression, disease interaction and medical interventions as well as for the support of the intelligent tutor agent in its task of helping the user train in the best clinical practices.

The MVP project places significant demands on physician-informants to render complex, multi-scale knowledge in a form that can be implemented computationally – naturally, with a knowledge engineer mediating between physicians and programmers. Physicians must distill their extensive and tightly coupled physiological and clinical knowledge into the most relevant subset, and express it in the most concrete terms. Not infrequently, they

are also called upon to hypothesize about the unknowable, like the state of a patient experiencing a pre-clinical stage of disease, or the state of a patient after an effective treatment that is never, in real life, followed up by objective tests. Such hypotheses reflect the mental models of specific experts, which might differ in subtle ways from those of other experts. However, such differences, we would suggest, have little bearing on the ultimate goal of this specific project: to create MVPs whose behavior is sufficiently life-like to further specific educational goals.

The process of extracting knowledge from the experts and from other sources – medical literature, existing repositories of structured medical knowledge – is at present not automated and requires the participation of a knowledge engineer. We have developed support tools for this task, but we do not expect to be able to automate this task any time soon. Knowledge models extracted from the experts and other sources are formulated in the metalanguage used by the OntoSem environment, specifically, its ontology.

The Ontology

The term "ontology" has been used of late to refer to a heterogeneous group of entities, so we will define our use of the term by referring to our ontology, the OntoSem ontology (stemming from the theory of Ontological Semantics; Nirenburg and Raskin 2004). The OntoSem ontology is fundamentally different from most other "ontologies" in its emphasis on rich property-based descriptions that are not present in the many hierarchical trees of words or concepts available both within the medical domain (e.g., UMLS; Bodenreider 2004) and outside of it (e.g., WordNet; http://wordnet.princeton.edu). One ontological model that does contain useful properties is the Foundational Model of Anatomy (Rosse and Mejino 2004; http://fma.biostr.washington.edu), which provides both inheritance (is-a) and merynomic (part-of) trees for elements of human anatomy. Concepts are linked using a mid-sized inventory of properties. In augmenting the OntoSem ontology for use in the medical domain, we have followed the FMA model in certain ways (e.g., with regard to naming conventions) in order to keep our knowledge resources compatible with what we believe will become the accepted standard. However, it would be incorrect to assume that FMA has answered all our needs in the medical domain: whereas it treats only anatomical objects, we need as thorough a treatment of relevant events and their relationship to objects, both anatomical and extra-anatomical.

The current OntoSem ontology contains around 9,000 concepts (objects, events and properties) that are connected through inheritance in a directed acyclic graph and are described by an average of 16 properties each, which can be inherited or locally defined. Most of the concepts belong to the general domain, apart from significant medical subgraphs. The ontology is written in a metalanguage whose atoms resemble English words and phrases (for ease of use by knowledge acquirers) but the semantics of these atoms is distinct from that of the (typically, ambiguous) English words with which they are homographous. When OntoSem is used for text processing applications, ontologically linked lexicons moderate between the text and the ontology. Various versions of the

OntoSem ontology have supported NLP applications in a half dozen languages for which lexicons of up to 50,000 words were compiled.

The inventory of basic primitive ontological properties—attributes and relations—numbers in the hundreds and is growing (though at a slower rate than the overall number of concepts). Domains and ranges are specified for each property. The ontology provides for multivalued selectional restrictions: *sem* for basic semantic constraints, *default* for stronger constraints, *relaxable-to* for the weakest acceptable constraints, and *value* for rigid constraints (*value* is used very rarely in the ontology: it primarily reflects actual properties of concept instances in the VP's long-term memory of assertions, see below).

Within the OntoSem environment, the ontology contains only generalized concepts, not instances of concepts; instances are stored in a fact repository, a model of the agent's long term memory of assertions. This knowledge base uses the same metalanguage as the ontology and is linked to the ontology but is conceptually and formally distinct. Just as the relationships between *types* of objects and events are stored in the ontology (the descriptive component of the knowledge base), the relationships between *instances* of objects and events are stored in the fact repository.

A cornerstone of creating a realistic virtual patient environment is providing for wide variation among instances of patients with a given disease. The basic, ontological model of a disease includes all relevant tracks (i.e., paths of progression), and each track provides many choice points that differentiate cases. Likewise, the basic, ontological model of a human includes all relevant properties of a human with all possible values, from eye color to genetic predisposition to esophageal cancer, to the reaction to all medications and procedures that can be administered through the system. The ontological models of both diseases and humans include default values for all properties, which permits the rapid authoring of patient instances with a focus on the property values that are actually important to the given simulation. For example, in none of the diseases modeled so far does eye color play a diagnostic role; however, every virtual patient instance must have some eye color, so an eye color is automatically attributed to him/her. (Note that if the user *did* ask what a VP's eye color was – once natural language support has been included – the patient could answer, something that might be important for later modeling of diseases like Alzheimer's, where memory tests play a diagnostic role.)

Proto-instances[12] of VPs are data sets for which specific values have been selected (explicitly or automatically) for all human and disease-related properties. A proto-instance is fully prepared to participate in one or more simulations, either by different users or by the same user who seeks to understand how patient outcome will differ with different interventions. A pedagogical utility of proto-instances is that each one can encapsulate a different teaching goal: e.g., "an achalasia patient for whom no treatment options are definitive" or "a GERD patient who will progress to adenocarcinoma if left untreated".

[12] The Knowledge Machine (Clark and Porter, nd) group uses this term with a similar meaning.

An *actual* instance is a proto-instance that is participating, or has participated, in a simulation. For each actual instance, a fact repository is dynamically generated that includes the patient's actual disease state, interventions, etc., over time.

Crucially for the VP project, OntoSem supports the encoding of complex events, also known as scripts, which represent typical sequences of events and their causal and other relationships. Scripts represent how individual events hold well-defined places in routine, typical sequences of events that happen in the world, with a well-specified set of objects that fill the different roles throughout that sequence. For example, if the sequence of events describes a colonoscopy, the participants will include the physician carrying out the procedure, the patient, and any number of other medical personnel; the tools will include the colonosope, various monitors and anesthesia; other props will include the operating table and medical gloves and gowns; events will include anesthetizing the patient, carrying out various procedures with the colonoscope; and so forth. Scripts can contain subscripts (e.g., the scripts of prepping and anesthetizing a patient), and they can be more or less fine-grained, depending on their intended use.

About Scripts

Although many types of knowledge can be represented using the simple slot-filler frames typical of ontologies, applications that rely on high-level reasoning also require scripts, which are descriptions of typical sequences of events, their causal and temporal relationships, and the objects that fill different roles throughout those sequences. Scripts have long been understood as necessary for high-level NLP and AI (e.g., Schank and Abelson 1977) but have been little pursued in practical system building due to the opinion, which we do not share, that the requisite knowledge acquisition is too expensive. Within OntoSem, scripts and simple slot-filler knowledge reside side by side in the ontology. They are employed both in NLP and in simulation/tutoring for essentially the same purpose: to support automatic reasoning.

Scripts represent complex events. There are at least two benefits of including scripts in an ontology: saving time by encoding information once then using across knowledge structures and applications, and giving an automatic reasoner sufficient knowledge upon which to base "intelligent" decisions. Consider the kinds of information that a human would want to be able to assume, rather than have to assert, when writing a script about swallowing: the tongue, larynx, pharynx, esophagus, lower esophageal sphincter and stomach (as well as hundreds of other objects) are all body parts of the same person; all of these are composed of tissue, which is composed of cells, some of which contain a nucleus; tissues are permeated by nerve cells of various kinds; the stomach is distal to the lower esophageal sphincter, which is distal to the body of the esophagus; when a nerve fires it sends a message to the brain… and so on. Once this information is ontologically encoded, a simulator can use it to reason about objects and events in scripts. For example, all variables in all scripts need to be bound; however, the OntoSem simulator can carry out unambiguous variable bindings automatically based on the meronymic relations encoded in the ontology, which turns out to be a significant savings in acquisition time.

We introduce a discussion of issues in developing scripts for simulation using an example. Below is a short excerpt from the script describing swallowing in humans. SWALLOW is the ontological concept that heads the swallowing script: the rest of the script is encoded as the filler of the HAS-EVENT-AS-PART property of SWALLOW as an hierarchical structure describing first the two main subevents of swallowing (OROPHARYNGEAL-PHASE-OF-SWALLOWING and ESOPHAGEAL-PHASE-OF-SWALLOWING), then further expanding those subevents.

The AGENT property of SWALLOW is constrained to HUMAN and the THEME to a BOLUS, which is a small mass of liquid or chewed solid food ready to be swallowed. The PRECONDITION for SWALLOW is that the BOLUS be located in the MOUTH.

The first complex subevent of SWALLOW is the OROPHARYNGEAL-PHASE-OF-SWALLOWING, which moves the BOLUS from the MOUTH to the LARYNX and is caused by a primarily agentive action (by contrast, the ESOPHAGEAL-PHASE-OF-SWALLOWING is unagentive). It has five top-level events: two MOTION-EVENTs, two RELAX-MUSCLE events and one CONTRACT-MUSCLE event (these have subevents as well, like various nerves firing, but we do not detail them here). We assume that the reader can follow the straightforward formalism with the help of the comments presented after semicolons.[13]

```
(SWALLOW                                    ;; the head of the script
  (AGENT              HUMAN)
  (THEME        BOLUS)
 (DURATION    10 (DEFAULT-MEASURE  SECOND))
  (PRECONDITION
     (LOCATION
        (DOMAIN           BOLUS)
        (RANGE            MOUTH)))
  (HAS-EVENT-AS-PART
     OROPHARYNGEAL-PHASE-OF-SWALLOWING      ;; expanded below
     ESOPHAGEAL-PHASE-OF-SWALLOWING))       ;; not shown here

(OROPHARYNGEAL-PHASE-OF-SWALLOWING
   (AGENT  HUMAN)
   (THEME  BOLUS)
  (DURATION    1 (DEFAULT-MEASURE  SECOND))
   (HAS-EVENT-AS-PART
     MOTION-EVENT=1            ;; bolus: from mouth to pharynx
     CONTRACT-MUSCLE           ;; contract pharynx, epiglottis closes
     MOTION-EVENT=2            ;; bolus: from pharynx to larynx
     RELAX-MUSCLE=1  ;; cricopharyngeus relaxes
     RELAX-MUSCLE=2))          ;; LES relaxes
```

[13] Our simulation program and NLP processors use exactly this format as input, obviating the need for multiple levels of knowledge representation, one for acquirers and another for machine processing. Currently, knowledge acquisition is a joint effort by a knowledge engineer and subject-matter experts, but we are working on developing a largely machine-initiative knowledge-elicitation system for script writing, relying on the methodologies developed for the Boas KE system (McShane et al. 2003).

```
(MOTION-EVENT=1                ;; bolus: from mouth to pharynx
   (AGENT       HUMAN)
   (THEME        BOLUS)
   (INSTRUMENT          TONGUE)
   (SOURCE      MOUTH)
   (DESTINATION         PHARYNX)
   (DURATION  .3 (DEFAULT-MEASURE  SECOND))
   (EFFECT
      (LOCATION
         (DOMAIN          BOLUS)
         (RANGE           PHARYNX))))

(CONTRACT-MUSCLE               ;; contract pharynx, epiglottis closes
   (AGENT                  HUMAN)
   (THEME                  SET-OF-CONSTRICTOR-MUSCLES-OF-PHARYNX)
   (DURATION             .5 (DEFAULT-MEASURE  SECOND))

   (EFFECT
     (OPENNESS
       (DOMAIN            EPIGLOTTIS)
       (RANGE  0))))

(MOTION-EVENT=2      ;; bolus from pharynx to larynx
   (AGENT              HUMAN)
   (THEME                BOLUS)
   (INSTRUMENT    SET-OF-CONSTRICTOR-MUSCLES-OF-PHARYNX)
   (SOURCE            PHARYNX)
   (DESTINATION    LARYNX)
   (DURATION  .5 (DEFAULT-MEASURE  SECOND))
   (EFFECT
      (OPENNESS
         (DOMAIN          EPIGLOTTIS)
         (RANGE           1))
      (LOCATION
         (DOMAIN          BOLUS)
         (RANGE           LARYNX))))

(RELAX-MUSCLE =1     ;; cricopharyngeus relaxes
   (AGENT              HUMAN)
   (THEME              CRICOPHARYNGEUS)
   (DURATION  1 (DEFAULT-MEASURE  SECOND))
   (EFFECT
      (PRESSURE
           (DOMAIN        CRICOPHARYNGEUS)
          (RANGE          < 5)))
```

```
(RELAX-MUSCLE=2                ;; LES relaxes
   (THEME       LES[14])
   (DURATION  9 (DEFAULT-MEASURE  SECOND))
   (EFFECT
     (PRESSURE
        (DOMAIN        LES)
        (RANGE          < 5))))
```

Component events in scripts also exist as free-standing concepts in the ontology. However, when used inside a script, such events become much more concrete than the corresponding general concepts: they accept specific fillers of many of their properties instead of the more generic constraints typical for general ontological concepts. For example, the THEME of either of the MOTION-EVENTs in the script excerpt above is restricted to BOLUS, while in the general ontology MOTION-EVENT has a much more generic constraint of PHYSICAL-OBJECT on its THEME property. Thus, components of scripts, though appearing in the ontology, resemble concept instances. Conceptually, they lie somewhere between concepts and actual, real-world instances and are therefore called ontological instances.

Scripts contain ontological instances of both events and objects (e.g., BOLUS in the script is, in fact, an ontological instance of the generic concept BOLUS). Basic ontological concepts provide the necessary information from which the script processor infers variable bindings across the component events of a script (e.g., that both motion events have *the same* BOLUS as their theme).In cases of ambiguity, ontological instances of events and are numbered.

The processor also infers meronymic relationships among ontological object instances. As humans reading the script excerpt above, we probably don't notice that we assume that the TONGUE is located in the MOUTH that is a body part of the HUMAN who is doing the swallowing. Nothing in the script explicitly says that these body parts must be bound to our HUMAN, but in fact they must be interpreted as such if the script is to be understood correctly and support a functional simulation. Requisite anatomical knowledge is recorded in the base ontology and is used by the script processor to infer bindings. For example, when the script processor encounters MOUTH in the SWALLOW script, it finds the shortest ontological path between MOUTH and each of the active ontological instances ("active" instances include any input thematic roles as well as any active instances in the script that called the current script via a HAS-EVENT-AS-PART). In this case, the only two active ontological instances would be the HUMAN and BOLUS instances that are the thematic roles input to the SWALLOW event. For a particular SWALLOW event, perhaps HUMAN-FR228 and BOLUS-22 are involved. The processor will determine that MOUTH has a PART-OF relationship to HUMAN, which is semantically closer than any relationship found between BOLUS and MOUTH. Thus, the processor will infer that the MOUTH in question is the one that is PART-OF HUMAN-FR228.

In cases where bindings can be unambiguously inferred from basic ontological information in this manner, script acquirers need not explicitly indicate them. However, in cases where bindings cannot be unambiguously inferred by the script processor, binding must be overtly indicated. For example, in a later part of the SWALLOW script we need to refer to various STRETCH-RECEPTORs and MOTOR-ENDPLATEs that are located in various body parts. However, there are huge numbers of STRETCH-RECEPTORs and MOTOR-ENDPLATEs throughout the body and, even if we created numerically differentiated ontological instances of them in the script that would not be sufficient to indicate which anatomical structure each was connected to.[15] In such cases, we must explicitly create the bindings in a "bind-variables" field of the given sub-event of the script, as shown below:

```
(PERISTALSIS                ;; moves BOLUS from LARYNX to
  (BIND-VARIABLES                ;; ESOPHAGEAL-SEGMENT-1
    (STRETCH-RECEPTOR (PART-OF-OBJECT  LARYNX))
    (MOTOR-ENDPLATE=1  (PART-OF-OBJECT  LARYNX))
    (MOTOR-ENDPLATE=2 (PART-OF-OBJECT  ESOPHAGEAL-SEGMENT-1)))
…
```

To reiterate, variable bindings are necessary for many script elements to support the complete and proper interpretation of scripts. However, in those cases when bindings can be unambiguously inferred based on core ontological properties, these bindings need not be overtly written by the script acquirer. Knowing when the script processor will and will not be able to make the correct bindings requires a good understanding of the basic ontological substrate.

The script excerpt above illustrated some basic properties of scripts, like the use of ontological instances, the possibility of embedding sub-events to any depth, the necessity of implicitly or explicitly creating variable bindings, and the anchoring of scripts in the basic ontology with the benefit of being able to rely on properties and values recorded outside of any given script. However, the expressive means described so far represent only a subset of those necessary to support an application like simulation. In this section, we describe a number of others.

Loops

It is frequent in scripts for a given type of event to occur many times with different property values. When such an event is, itself, complex (having subevents as its parts), the desire for economy suggests the need for loops: calls for the same type of complex event with argument values specified for each separate ontological instance. The event of PERISTALSIS, by which muscles of the esophagus (involuntarily) push food through the esophagus to the stomach, is a case in point. According to our conceptual model, the esophagus contains twelve segments whose names correlate with vertebral segments (C6, C7 and T1-T10). The peristalsis events that take the bolus from the larynx to C6, and

[15] We certainly do not want to create a separate concept for each stretch receptor in the human body, differentiated only by which body part they attach to, since such concepts would number in the millions.

from T10 to the stomach, have special properties and are handled separately; however the peristalsis events that take the bolus from C7→T1, T1→T2… T9→10 are essentially the same. We use a special notation to refer to script elements that are part of loops: prefixing them by an asterisk. An excerpt from the ESOPHAGEAL-PHASE-OF-SWALLOWING that shows this convention is as follows:

```
(ESOPHAGEAL-PHASE-OF-SWALLOWING
  (AGENT   *nothing*)          ;; the process is involuntary
  (THEME    BOLUS)
   (DURATION  8 (DEFAULT-MEASURE  SECOND))
  (HAS-EVENT-AS-PART
     …
     *PERISTALSIS
        (SOURCE         C6-SEGMENT-OF-ESOPHAGUS)
        (DESTINATION C7-SEGMENT-OF-ESOPHAGUS)
      *PERISTALSIS
        (SOURCE                  C7-SEGMENT-OF-ESOPHAGUS)
        (DESTINATION T1-SEGMENT-OF-ESOPHAGUS)
    …
     *PERISTALSIS
        (SOURCE                  T9-SEGMENT-OF-ESOPHAGUS)
        (DESTINATION T10-SEGMENT-OF-ESOPHAGUS)
    …  ))
```

When the event *PERISTALSIS is expanded into its subevents, the ontological instances are also referred to using the asterisk notation to show that, during the running of the script, they will be turned into distinct instances of the given concepts:

```
(*peristalsis
  (source        *segment-of-esophagus)
  (destination  *segment-of-esophagus)

  (bind-variables
    (*stretch-receptor
          (part-of-object        *source)
     (*motor-endplate=1
        (part-of-object           *source))
     (*motor-endplate=2
          (part-of-object         *destination))
      …)      ;; more variable bindings

  (has-event-as-part  ;; subevents of *peristalsis
        *stretch
        *fire-nerve
        *stimulate=1
        *stimulate=2
```

```
        *contract-muscle
        *relax-muscle
        *motion-event
        *relax-muscle=2)
  …   )
```

To reiterate, special expressive means are needed for loops to permit economy of knowledge representation while still permitting the script interpreter and simulation program to correctly track variables.

Conditionals

A scripting language must support the use of conditionals, an expressive means that we employed frequently in our modeling of the function and diseases of the esophagus. An example is gastric pressure: each person has a basic gastric pressure, which is within a range of possible values, and this basic pressure is affected by gravity (whether a person is erect or supine) and the amount of food in the person's stomach. If the person is upright, there is no added pressure from gravity, but if he is supine then 5-10 torr is added. If the person has a large amount of food in the stomach (represented as greater than .7 on an abstract scale of quantity) then 7-10 torr is added, and if he has less food in the stomach then less pressure is added. This information is recorded in scripts as follows:

```
(stomach
  (pressure (sum      (pressure-basic-gastric
                      pressure-from-gravity
                      pressure-from-food)))[16]
  (pressure-basic-gastric (<> 5 10))
  (pressure-from-gravity
    (if
      (spatial-orientation
         (domain      human)
         (range vertical))
      then  0)
    (if
       (spatial-orientation
          (domain     human)
          (range      horizontal))
      then  (<> 5 10))))
  (pressure-from-food
    (if
       (location
         (domain      stomach)
         (range       (food (quant  (> .7)))))
      then  (<> 7 10))
   (if
```

[16] Abdominal pressure is another variable but we omit it in this exposition.

```
        (location
            (domain        stomach)
            (range (food   (quant  (<>  .3 .6)))))
     then  (<> 4  6.9))
    (if
        (location
            (domain        stomach)
            (range (food (quant  (< .3)))))
     then  (< 4)))))
```

Within larger scripts of esophageal function, the various values for gastric pressure are combined with values for the pressure of the LES (lower esophageal sphincter) to determine whether or not a person will experience reflux. If he does, then he will experience pain, with the duration of the reflux affecting the severity of pain. If the reflux continues over a long period of time, the esophagus will enter a disease state in which its actual properties change permanently. If the disease is treated (through medications) at an early stage, the esophagus will heal to an extent but if treated at a later stage, healing will be less complete or impossible; and so on. All of these conditionals, and many more, are included in our current swallowing script, which has been implemented as a simulation.

Time

Time is an essential element of scripts, especially if they are used for simulation. In the case of swallowing, the entire process takes around 10 seconds, with different subscripts accounting for different portions of that, as indicated by the DURATION slot.

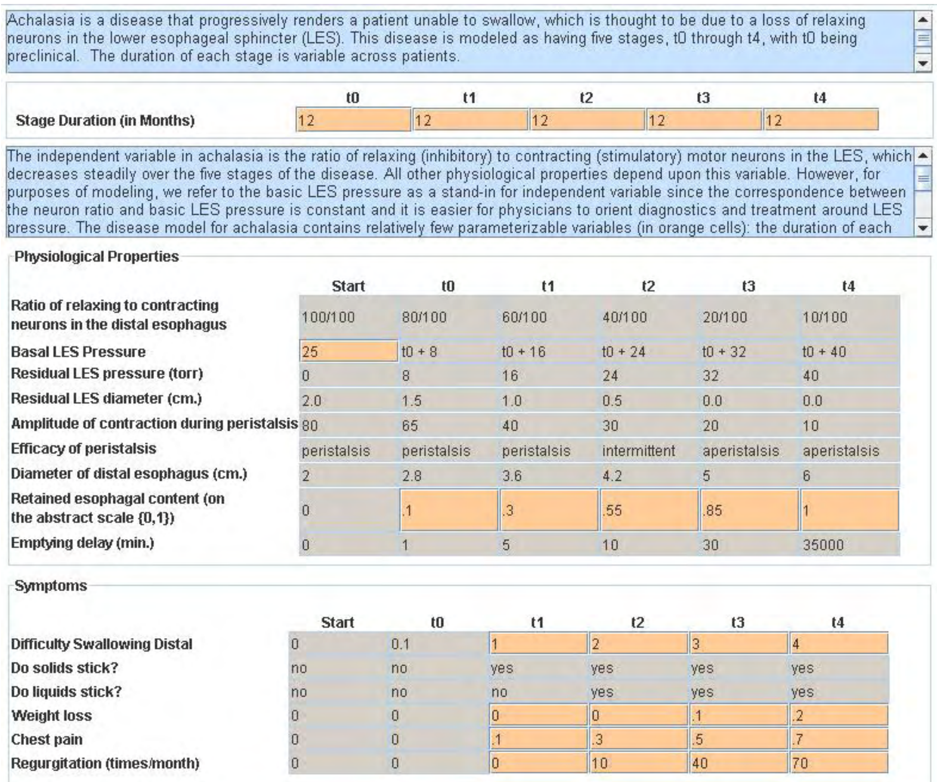
In medical simulation, modeling time is necessary in order to account for the progression of disease. An example is the growth of a tumor. Once a tumor is assigned a growth rate, it will grow at that rate when the script is run. Physiological effects of the tumor at different sizes are anticipated in the script. For example, a small tumor in the esophagus will show no symptoms, a medium-sized tumor will lead to mild dysphagia (difficulty swallowing), a large tumor will block solids but still permit liquids to go down, and a massive tumor will block all substances from passing. The notions of small, medium and large are determined by the physicians acting as subject-matter experts during script development.

Patient authoring

In addition to acquiring general medical knowledge, the project also requires the creation of an extensive library of actual virtual patients that exhibit a variety of diseases and genetic and cognitive traits that influence disease progression and response to treatment. We have developed an advanced interactive environment to support patient authoring by physicians who may not know about the internal workings of our models of disease and treatment.

The patient creation process for all diseases begins with providing basic information about the patient: name, age, gender, weight, race, etc. We omit this aspect of the

interfaces, as well as other aspects that are easily described in prose, in the screen shots below for reasons of space.


Achalasia is a disease that progressively renders a patient unable to swallow, which is thought to be due to a loss of relaxing neurons in the lower esophageal sphincter (LES). This disease is modeled as having five stages, t0 through t4, with t0 being preclinical. The duration of each stage is variable across patients.

	t0	t1	t2	t3	t4
Stage Duration (in Months)	12	12	12	12	12

The independent variable in achalasia is the ratio of relaxing (inhibitory) to contracting (stimulatory) motor neurons in the LES, which decreases steadily over the five stages of the disease. All other physiological properties depend upon this variable. However, for purposes of modeling, we refer to the basic LES pressure as a stand-in for independent variable since the correspondence between the neuron ratio and basic LES pressure is constant and it is easier for physicians to orient diagnostics and treatment around LES pressure. The disease model for achalasia contains relatively few parameterizable variables (in orange cells): the duration of each

Physiological Properties

	Start	t0	t1	t2	t3	t4
Ratio of relaxing to contracting neurons in the distal esophagus	100/100	80/100	60/100	40/100	20/100	10/100
Basal LES Pressure	25	t0 + 8	t0 + 16	t0 + 24	t0 + 32	t0 + 40
Residual LES pressure (torr)	0	8	16	24	32	40
Residual LES diameter (cm.)	2.0	1.5	1.0	0.5	0.0	0.0
Amplitude of contraction during peristalsis	80	65	40	30	20	10
Efficacy of peristalsis	peristalsis	peristalsis	peristalsis	intermittent	aperistalsis	aperistalsis
Diameter of distal esophagus (cm.)	2	2.8	3.6	4.2	5	6
Retained esophagal content (on the abstract scale {0,1})	0	.1	.3	.55	.85	1
Emptying delay (min.)	0	1	5	10	30	35000

Symptoms

	Start	t0	t1	t2	t3	t4
Difficulty Swallowing Distal	0	0.1	1	2	3	4
Do solids stick?	no	no	yes	yes	yes	yes
Do liquids stick?	no	no	no	yes	yes	yes
Weight loss	0	0	0	0	.1	.2
Chest pain	0	0	.1	.3	.5	.7
Regurgitation (times/month)	0	0	0	10	40	70

Figure 13. An excerpt from the authoring interface for achalasia.

Disease models break down into two major classes based on whether or not the physiological causal chains underlying the disease are well understood. In cases where physiological causal chains are relatively poorly understood – as for achalasia, scleroderma esophagus and Zenker's diverticulum – the simulation is primarily driven by temporal causal chains. Each disease is divided into conceptual stages, with each stage being associated with clinically observed physiological changes and symptom profiles. As simulated time passes, the patient's state changes incrementally, calculated using an interpolation function that incorporates the start value of each property at the beginning of the disease and the end value for each conceptual stage. Figure 13 shows these aspects of the model of achalasia, as presented to patient authors. The text in the blue background explains each aspect of the model using methods of progressive disclosure (a small text field with a scroll bar), which permits users with different levels of experience using the system to use the same interface without the explanatory materials becoming cumbersome. The explanatory text conveys important aspects of how the recorded property values are interpreted within the model and processed by the simulation engine, like which property values impaired by a disease can be reversed given an effective treatment, and which variables are independent and which are dependent. In short, the knowledge used by the simulator goes well beyond what is needed to parameterize a new patient instance, and it is conveyed to patient authors as text in order to clarify – albeit in encapsulated form – how the instantiated model works.

The gray cells indicate values that are fixed for all patients, since permitting their variation is not necessary for either of our immediate goals: (a) generating automatic function in the simulation (e.g., if a given biological pathway can be affected by medication, then it must be parameterizable) and (b) permitting noteworthy variation among patients within a teaching context. The orange cells indicate property values that can be changed for each patient, within ranges visible by mousing over the given cell. This division between paramaterizable and non-parameterizable property values points up an important benefit of making our models accessible to the community: for the current teaching application, it was appropriate to make certain property values parameterizable within certain ranges; however, for some other application it might be necessary to make more of these values more variable across patients, which can be readily done with no changes required of the simulation engine.

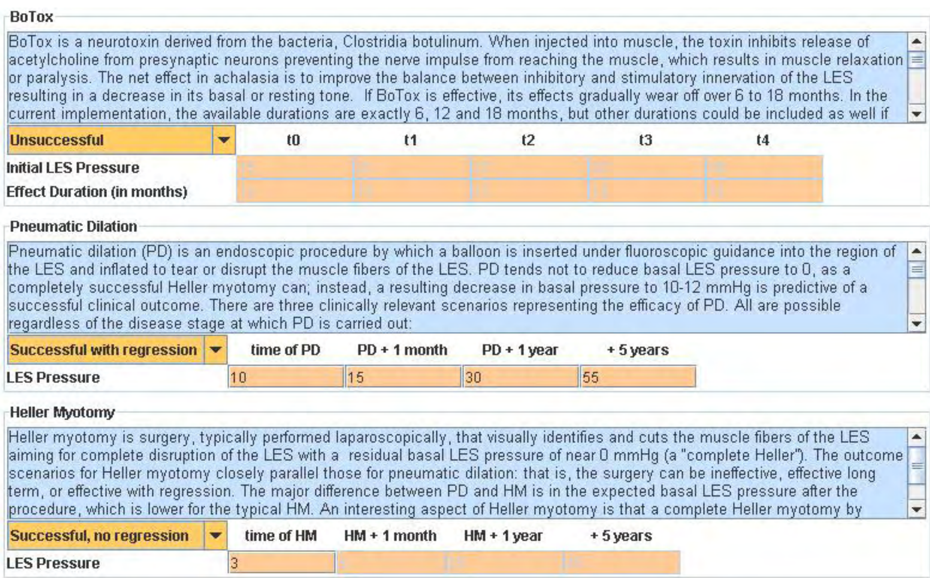


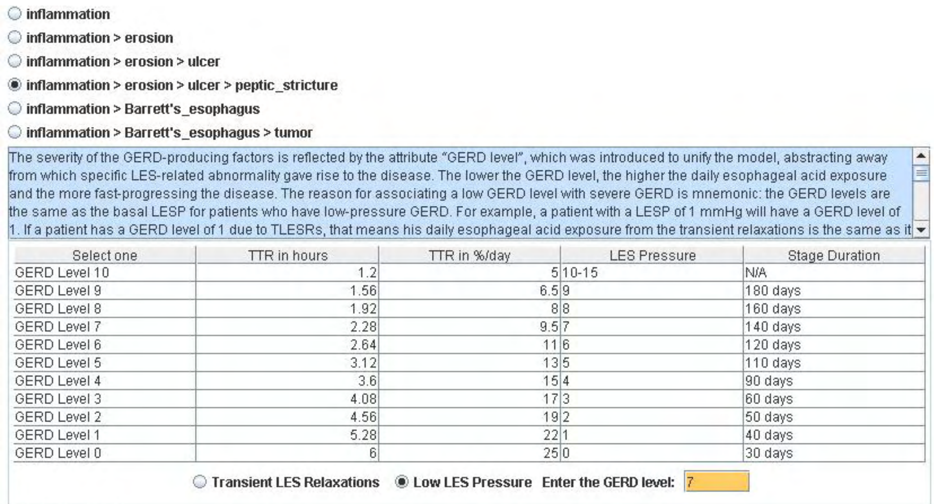
Figure 14. Treatment outcomes for achalasia patients.

The remaining aspect of patient parameterization for achalasia regards treatments. There are three treatment options, each explained in the associated blue shaded text fields (see Figure 14). Each has three potential options: unsuccessful, successful with regression, and successful without regression. If a treatment is unsuccessful, as for BoTox in Figure 14, there are no further choices to be made: the patient's condition is unchanged (note that the values in the cells of the corresponding table are grayed out). If a treatment is successful with regression (as for pneumatic dilation in Figure 14), the author must choose the rate of regression of the basal pressure of the lower esophageal sphincter (LESP) over time: its value one month, one year and five years after the procedure. (Although LESP is actually dependent upon the ratio of contracting to relaxing neurons, it is conceptually easier for clinicians to reason using LESP). After a procedure, most physiological properties and symptoms retain the original correspondences with LESP shown in the tables in Figure 13; however, the efficacy of peristalsis and diameter of the distal esophagus never improve once compromised, as explained in the blue text field. If a treatment is successful without regression (as for Heller myotomy in Figure 14), only

the original post-procedure LESP must be indicated, with most other property values following suit, as described above.
The other class of diseases modeled in the system are those for which physiological causal chains are quite well understood. GERD, LERD and LERD-GERD are all of this type. For reasons of space, we highlight just one aspect of the causal modeling of GERD and how it is reflected in the patient authoring interface (see McShane et al. 2007a,b for more in-depth descriptions of these disease models).

GERD can be defined as any symptomatic clinical condition that results from the reflux of stomach or duodenal contents into the esophagus. Based on a person's inherent predispositions (no biomarkers have yet been discovered), the disease can take one of six paths, shown at the top of Figure 15. The author selects one path for his patient, which sets associated property values in the patient. The two sources of GERD are abnormally low pressure of the lower esophageal sphincter (LES) (< 10 mmHg), or an abnormally large number or duration of transient relaxations of the LES (TLESRs), both of which result in increased acid exposure of the esophageal lining. The text in blue in Figure 15 (which is quite long; note the slider size) describes how LESP and/or TLESRs are used as independent variables in the model. We repeat an excerpt from that text here as an example of how text complements the formal aspects of disease models:

> The severity of the GERD-producing factors is reflected by the attribute "GERD level", which was introduced to unify the model, abstracting away from which specific LES-related abnormality gave rise to the disease. The lower the GERD level, the higher the daily esophageal acid exposure and the more fast-progressing the disease. The reason for associating a low GERD level with severe GERD is mnemonic: the GERD levels are the same as the basal LESP for patients who have low-pressure GERD. For example, a patient with a LESP of 1 mmHg will have a GERD level of 1. If a patient has a GERD level of 1 due to TLESRs, that means his daily esophageal acid exposure from the transient relaxations is the same as it would have been if he had had a basal LESP of 1. Using GERD level as the anchor for modeling provides a simple mechanism for incorporating a patient's lifestyle habits into the simulation: whenever he is engaging in bad lifestyle habits (assuming he has GERD-related sensitivities to those habits), his GERD level decreases by 1. For patients with a baseline GERD level of 10 – which is not a disease state – this means that engaging in bad habits is sufficient to initiate GERD and discontinuing them is sufficient to promote healing without the need for medication. For patients with a baseline GERD level of less than 10, lifestyle improvements can slow disease progression but not achieve the healing of previous esophageal damage.



Select one	TTR in hours	TTR in %/day	LES Pressure	Stage Duration
GERD Level 10	1.2	5	10-15	N/A
GERD Level 9	1.56	6.5	9	180 days
GERD Level 8	1.92	8	8	160 days
GERD Level 7	2.28	9.5	7	140 days
GERD Level 6	2.64	11	6	120 days
GERD Level 5	3.12	13	5	110 days
GERD Level 4	3.6	15	4	90 days
GERD Level 3	4.08	17	3	60 days
GERD Level 2	4.56	19	2	50 days
GERD Level 1	5.28	22	1	40 days
GERD Level 0	6	25	0	30 days

Figure 15. An excerpt from the patient authoring process for GERD.

As is clear by the table, when the author selects the GERD Level (he chose 7 in Figure 15) the duration of each stage of the disease and the total time in reflux (TTR) are automatically selected for him. Other aspects of the patient authoring interface permit authors to select lifestyle habits for their patients, whether those lifestyle habits affect their GERD, their symptom profile and their response to medications. Our point in this example is to show that even when a disease is modeled using causal chains that are encoded in quite complex ontological scripts and realized in even more complex simulation programs, the conceptual substrate of the basic models can readily be shared with – and contributed to – by the larger community.

Discussion

The patient authoring interface in MVP highlights key aspects of the cognitive model of each disease, providing patient authors with explanations of the choice space without either repeating all the information about each disease available in textbooks or expounding upon the implementation of the simulation engine. The core aspects of each disease model include which property values are parameterizable among patients and which ones are fixed for all patients, what ranges of values are permitted for each property at each stage of the disease (seen by rolling over cells in the interface), how "healing" is interpreted with respect to each property value affected by a disease, and how parameterizable property values are used to "bridge" unknown aspects of diseases, like as yet undiscovered genetic influences. The grain-size of description – including which aspects are made parameterizable and which physiological causal chains are included in the model – is influenced by the given application but could easily be changed to suit other applications using our ontologically grounded knowledge encoding methodology. Let us consider this last point in more detail using the example of GERD. As shown in Figure 15, by selecting a GERD level, the author automatically sets the duration of each conceptual stage of the disease and the total time in reflux per day. For a pedagogical application, these fixed correspondences are very useful. However, we plan to use this knowledge environment for other applications as well, like automatically analyzing electronic patient records both to validate the model and to learn new population-level clinical knowledge. It is likely that some patients fall outside the range

of expected outcomes of our current model, which can lead in two directions: either expanding the current model by making more aspects parameterizable, or creating a second, non-pedagogical version of GERD, thus permitting the pedagogical version to retain strong correspondences that are useful as a conceptual architecture, abstracting away from confounding cases. In fact, our knowledge environment can accommodate any number of versions of a disease model suited to different applications. Similarly, the mentoring module (which we did not discuss here) can also accommodate any number of versions: currently, our virtual mentor reflects one set of clinical preferences, but there could be a entire population of virtual mentors reflecting variations on clinical management practices. Recording disease and mentoring models using ontological scripts (rather than, for example, very large rule sets) permits such variation to be readily recorded and managed.

We have designed our knowledge environment such that we can readily collaborate with the broader community. For example, as more genetic influences on diseases are discovered, and more causal chains are understood, these will be used to replace temporal causal chains with physiological ones. We have made our disease models inspectable so that not only can experts assess them in terms of how virtual patients behave in a simulation, but also in terms of the core tenets of the mental models of the physicians who contributed to their development.

Language Processing

The OntoSem environment that provides the knowledge substrate for the MVP project has been first developed to support natural language processing. So, it is not surprising that OntoSem supports NLP in the MVP project, too. In this section, we will very briefly describe the NLP tasks within the MVP project.

Figure 16 shows the architecture of the OntoSem text understanding system (e.g., Nirenburg and Raskin 2004; Beale et al. 1995, 2003, 2004; McShane et al. 2005a, 2006). Text analysis involves preprocessing, syntactic analysis, semantic analysis, and discourse processing (including contextual compositional semantics for, among other things, reference resolution). The static resources used by the system include: (1) A property-rich, general-purpose ontology of about 9,000 concepts that contains information about the types of entities and events known to the system, together with their properties; (2) An ontologically-linked lexicon of about 20,000 words and phrases that includes syntactic and semantic information to support disambiguation; (3) An onomasticon, or lexicon of proper names; and (4) A repository of the given language processing agent’s long-term memory of assertions (LTM-A). Although ours is a general-purpose ontology, coverage of the medical domain has been expanded for the MVP application to include detailed descriptions of human anatomy, complex events (i.e., scripts) reflecting normal and pathological physiological processes (cf. Schank and Abelson 1977), and best clinical practices for the diagnosis and treatment of diseases.

The primary goal of OntoSem analysis is to automatically generate unambiguous text-meaning representations (TMRs) of input, represented in a metalanguage that is grounded in the OntoSem ontology. Semantic analysis begins with word sense disambiguation and

the establishment of basic semantic dependencies. This process relies on knowledge about selectional restrictions – mutual constraints – that is stored in the lexicon and ontology entries that are activated by the input text. In OntoSem, selectional restrictions are multi-valued, which allows for contextual tightening or relaxation of constraints when building TMRs. When successful, this process yields *basic* TMRs. However, no lexicon or ontology, no matter how broad and deep, can guarantee successful disambiguation of all inputs. This is why OntoSem incorporates a number of methods and algorithms for dealing with cases of residual ambiguity and zero output. These include a) an algorithm for the dynamic tightening and relaxing of selectional restrictions based on the properties of the input text; b) a statistically trained disambiguation algorithm based on comparing weighted distances between pairs of concepts in the ontology; c) an algorithm for the unidirectional application of selectional restrictions to process input words that are not listed in the system's lexicon; and d) an algorithm for deriving TMRs from fragments of input in cases of failure to produce a TMR for a complete sentence. The use of both knowledge-based and empirical methods demonstrates that OntoSem takes a practical, task-oriented approach to text analysis rather than a method-oriented one.

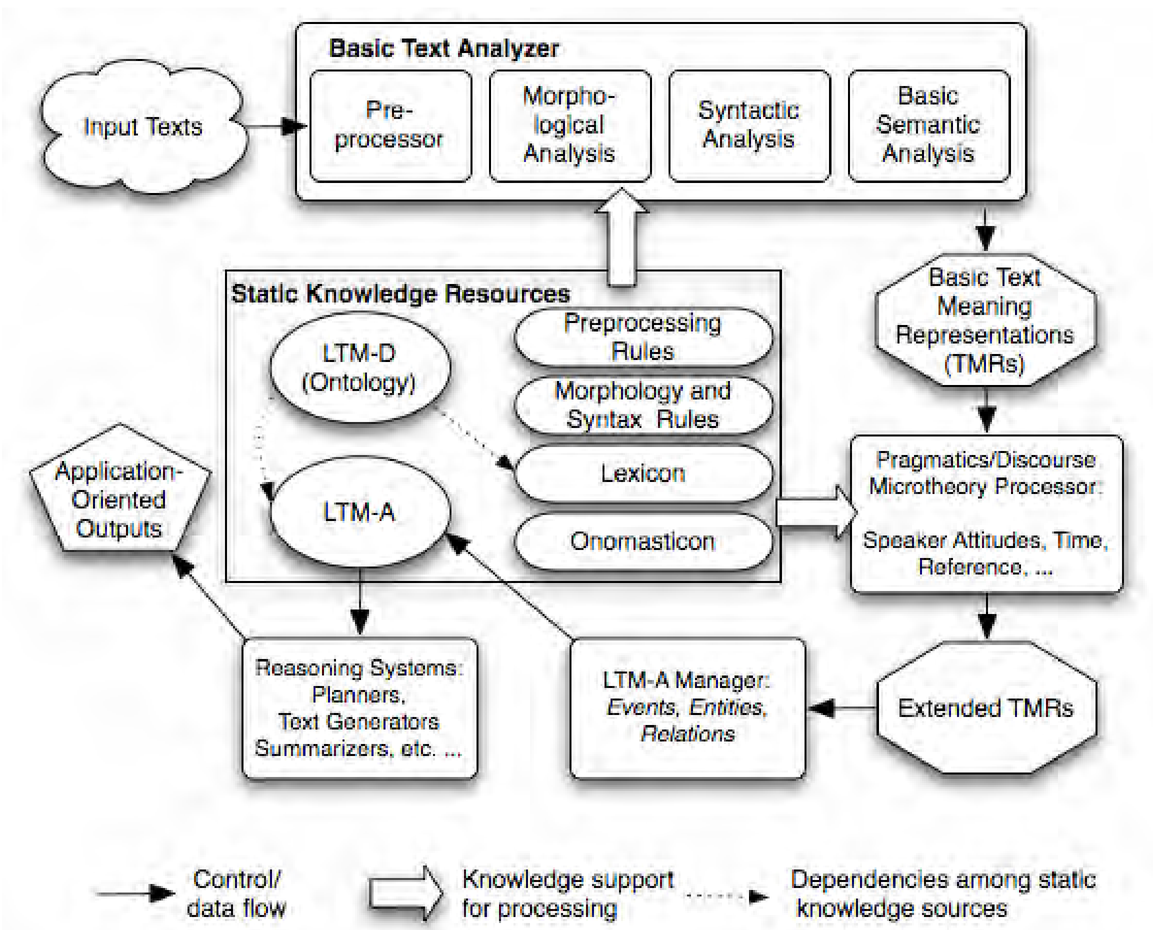


Figure 16. The architecture of the OntoSem text analyzer. Note the use of the long-term memory of descriptions (the ontology) and the long-term memory of assertions.

Once a basic TMR is produced, the analyzer attempts to generate an *extended* TMR by carrying out advanced aspects of text analysis, formulated as "microtheories." In the proposed project we seek to enhance the microtheories of reference resolution and unexpected input processing. Among other benefits, this will enhance the agent's ability to leverage information in text for the population of its LTM-A.

All development and evaluation in OntoSem is supported by DEKADE, our Development, Evaluation, Knowledge Acquisition and Demonstration Environment (McShane et al. 2005c). Among DEKADE's many functions are: accepting text input to be run through the analyzer; permitting the results of each stage of analysis (preprocessing, syntax, semantics, procedural semantics) to be viewed and edited using textual or graphical interfaces; offering easy access to the knowledge resources (lexicon, ontology, onomasticon and LTM-A) so that they can be edited and supplemented during work on a text or using other acquisition methodologies; and supporting formal evaluation of the resulting TMRs (see Nirenburg et al. 2004 for results of our first evaluation). The DEKADE environment, completed only recently, has drastically increased the speed of OntoSem system development and the efficiency of system evaluation.

OntoSem incorporates a number of earlier processors – the NMSU CRL TIPSTER tokenizer (Davis and Ogden 2000), the BBN IdentiFinder tokenizer (Bikel et al. 1999), the Bikel parser (Bikel 2002), the Hunter-Gatherer constraint satisfaction architecture (Beale 1997) and the OntoSearch algorithm for finding cheapest paths between two concepts in the OntoSem ontology (Onyshkevych 1997). A number of existing knowledge resources were used by OntoSem knowledge engineers to help them acquire the ontology, the lexicon and other knowledge resources. These include the XTAG grammar (www.cis.upenn.edu/~xtag/), the Penn Treebank (www.cis.upenn.edu/~treebank/), the Foundational Model of Anatomy (sig.biostr.washington.edu/projects/fm/) and WordNet (wordnet.princeton.edu).

NLP Dialog in MVP

The dialog-enhanced MVP system, MVP-D, is under development at the moment of writing. It will supplant the menu-based interaction between the user and the cognitive agents in the system – the virtual patient and the tutor. In MVP-D (as opposed to MVP), the VP will not have direct access to its physiological state. Instead, like a real person, it will perceive its own "body", including sensations of pain and other symptoms, as warranted. The VP will thus be able to reason about whether to consult a physician, whether to stick to a medication regimen, etc. The VP will also remember the gist of its prior conversations with the user (the attending physician) as well as the beliefs/facts about its physiology that it accumulates over time in its LTM-A. The communication with the user (the attending physician) will be conducted in natural language. Natural language will also be used for the communication between the tutor and the user. The tutor will have access to (i.e., memory of) all of the past dialog interactions in the system, the current state of the student's knowledge about the patient, and the knowledge base of best practices. The first version of MVP-D will operate using the current repertoire of the 6 esophageal diseases supported by MVP. In parallel to the proposed work, and under

different funding, we will continue work on acquiring knowledge about heart disease for the VP. Once such knowledge is acquired, MVP-D will be evaluated on an extended knowledge base and broadened dialog coverage.

Learning new concepts

A core component of MVP is the module supporting dialog between the VP and the student. One of the important aspects of creating natural-sounding dialog is to allow different VPs to have different ways of expressing themselves, including different degrees of knowledge about medicine and medical terminology. Students should be as adept at conducting an interview with an uninformed patient (in which case many paraphrases of medical terms might be needed) as with a fellow physician. Another important aspect of conducting an interview is educating the VP about his condition, his medication regime, needed lifestyle improvements, options for treatments, and so on. The result of all of this communication has to be learning on the part of the VP: after all, if the student chooses to explain to the VP that the feeling of having something stuck in one's throat is called a "globus sensation", we would expect the VP to remember that 2 minutes later (and, perhaps, at the next visit as well). At a minimum, the patient will eventually need to remember the name of his disease, his medications, and so on.

In developing the NLP support for MVP, dynamically adding to the VPs knowledge repository is of primary concern, and is a good example of an adaptive system module. That is, not only will the patient learn about its condition and treatment through verbal interactions, it will be able to put this knowledge to use in decision-making, one of the key functionalities of the cognitive agent. For reasons of space, in this paper we use a sample interaction to give an informal, content-oriented description of this adaptive process – a more formal analysis would have required a description of the various static and dynamic knowledge resources underlying the system and, specifically, its natural language processing component (this information can be found in [6], among others).

Suppose that during a patient interview the student asks the VP if the latter ever experiences regurgitation, and suppose that the VP does not understand the term *regurgitation* (that is, the entry for this word is absent from the VP's semantic lexicon).

Background: Each agent in the system is supplied with its own version of the knowledge resources available to the system. This means, for example, that the tutor knows much more about diseases and clinical practices than the virtual patient. The latter's lexicon and ontology are deliberately filtered to reflect an average lay person's knowledge of medicine. During patient authoring, the author selects the level of medical knowledge of the patient, and the lexicon and ontology supplied to the patient are populated accordingly.

The VP will ask for clarification by issuing a dialog turn such as: "What is regurgitation?" The goal of this subdialog is to learn the new term.

First let us consider the eventuality when the human user responds by suggesting a synonym for regurgitation. If the synonym is in the patient's lexicon, then the VP first

learns the new lexicon entry, whose semantics will be the same as for the known synonym, and then responds to the original question using this semantic interpretation. As a result of this process, the VP's lexical stock is increased, so that the next time the formerly unknown word is used in a dialog, there will be no need for the clarification subdialog. Of course, the student might provide a synonym that is not in the patient's lexicon, in which case the patient may opt to continue the clarification subdialog.

Instead of a synonym, the user may provide a description of what regurgitation is. On receiving this input, the patient's goal is to match the description to a concept in its ontology. If such a concept is found, then a new lexicon entry for the unknown word (in this case, regurgitation) is created, and the matching concept is used in the entry's semantic description. If no concept is a close enough match, then the VP must learn a new concept.

Suppose the student supplies the following definition of regurgitation: "The return of partially digested food from the stomach to the mouth."[17] And suppose the VP knows all the words in the above explanation.[18] The language analyzer processes this input and comes up with the following text meaning representation (only relevant information is presented; ontological concepts are shown in small caps):

RETURN-1 (which IS-A MOTION-EVENT)

THEME	INGESTIBLE-105
SOURCE	STOMACH-1
DESTINATION	MOUTH-1

The above text meaning representation contains numbered instances of ontological concepts as heads of frames and values of properties. But when comparing this structure with the concepts in the patient's ontology, we disregard instance numbers, in effect treating this text meaning representation as a candidate ontological concept. If a sufficiently close match is found in the VP's current ontology, then the putative new ontological concept is discarded and the already existing best match is used to describe the semantics of the unknown word. If the best match is not considered close enough, then the candidate concept is "promoted" to a regular concept in the ontology. In our example, the search in VP's "lay person" ontology for the best match of the above text meaning representation is the concept VOMIT:[19]

	VOMIT
IS-A	ANIMAL-SYMPTOM, MOTION-EVENT
THEME	INGESTIBLE
SOURCE	STOMACH
DESTINATION	MOUTH
VELOCITY	> .8

---

[17] This definition is a real one from The American Heritage Science Dictionary.

[18] This is actually the case with our language processing resources – the complete (unfiltered) English lexicon in our system covers over 30,000 word senses, and the ontology used to explain these senses consists of over 9,000 concepts, each of which has on average 16 properties defined for it.

[19] The process briefly described here is a special case of learning ontologies and lexicons by reading text and analyzing it using our OntoSem environment for meaning extraction. This work is described in more detail in [10].

In other words, the event most closely associated with food coming up that the patient knows about is vomiting. If the two concepts are judged sufficiently similar, the concept VOMIT will be used to describe the semantics of regurgitation. If not, the candidate concept will be included in the ontology and given the name REGURGITATION. If the results of clustering are uncertain, the VP may opt for a continuation of the clarification subdialog by passing the responsibility for this decision to the user, that is, by asking, e.g., “Do you mean vomiting?” If the user agrees that these are sufficiently close, that settles the issue. But if the user considers it important to distinguish between vomiting and regurgitating, then he or she will respond to the effect that regurgitation is like vomiting but not as forceful. In this case, the learning module in the VP will add the concept REGURGITATION to the ontology; this new concept will be similar to VOMIT but have the additional property that its VELOCITY is lower (the property VELOCITY will be licensed for VOMIT on account of its being an ontological descendant of MOTION-EVENT, for which the property of VELOCITY is defined).

It is important to stress that the “maximum coverage” ontology that our system can use already has the concept REGURGITATION, along with its siblings VOMIT and REFLUX. All of the above concepts are, in fact, children of BACKWARDS-MOTION-OF-INGESTED-SUBSTANCE, which itself is a child of both ANIMAL-SYMPTOM and MOTION-EVENT (exploiting multiple inheritance). The experiencer of all of these events is a MEDICAL-PATIENT, but the events differ with respect to five properties, as shown in the following table illustrating a subset of the knowledge in the “maximum-coverage ontology.”

	REGURGITATE	**VOMIT**	**REFLUX**
IS-A	BACKWARDS-MOTION-OF-INGESTED-SUBSTANCE		
THEME	BOLUS	CHYME	CHYME
SOURCE	ESOPHAGUS	STOMACH	STOMACH
DESTINATION	THROAT	MOUTH	ESOPHAGUS
VELOCITY	< .2	> .8	< .2
INSTRUMENT	-	MUSCLE-LAYER	-

The above underscores our commitment to adaptivity in the system. Indeed, we could have made the VP omniscient, at least in the domain of the diseases that it carries. Instead, we chose to model a much more realistic situation without “cheating” by allowing all agents to operate with full access to all knowledge at all times.

Utility Beyond the Current Application

The MVP project can be viewed as just one of a number of applications in the area of intelligent clinical systems. The latter, in turn, can be viewed as one of the possible domains in which one can apply modeling teams of intelligent agents featuring a combination of physical system simulation and cognitive processing.

So, in the most general terms, our work can be viewed as devoted to creating working models of societies of artificial intelligent agents that share a simulated “world” of an application domain with humans in order to jointly perform cognitive tasks that have until now been performed exclusively by humans. Sample applications of such models include:

- a team of medical professionals diagnosing and treating a patient (with humans playing the role of either a physician or a patient)
- a team of intelligence or business analysts collecting information, reasoning about it and generating analyses or recommendations (with humans playing the role of team leader)
- a team of engineers designing or operating a physical plant (with humans playing the role of team leader)
- a learning environment (where humans play the role of students).

As can be seen, this work is at the confluence of several lines of research – cognitive modeling, ontological engineering, reasoning systems, multi-agent systems, simulation and natural language processing.

A basic development and delivery environment for this work should consist of at least:

1. the working model of a society of intelligent agents;
2. a simulated physical system; and
3. an interface for communicating with the human members of the society.

The intelligent agents must be endowed with knowledge about

a) their world;
b) their own character traits;
c) a set of goal types;
d) a set of plans to achieve each of the goals; and
e) their experiences, in the form of memories of past actions and states of
  i) the physical system;
  ii) other agents in the environment; and
  iii) themselves.

In a typical application, these models must be fully or partially shared among the agents in the society. The agents are capable of

1. (simulated) perception of changes in the (simulated) physical system;
2. understanding communications from other agents (as part of natural language dialog)
3. reasoning about the state of the physical system, themselves and other agents;
4. deciding on what to do next on the basis of the state of the world and their own goals and plans;
5. performing (simulated) physical actions in the world;
6. performing communicative actions (as part of natural language dialog).

Natural language processing (NLP) is important in such an undertaking because, unlike the majority of multi-agent systems currently under development, our work always presupposes the participation of people, not only artificial intelligent agents. Natural language is the most natural way of communicating for people, even if having to use it

introduces an additional complexity to building agent systems. Our approach to NLP is knowledge-based and is fully compatible with our approach to general cognitive modeling, reasoning and problem solving. The meanings of natural language words and expressions are encoded in our approach using the same ontology that describes the general world model of the agents. The model of dialog we use adopts the same goal- and plan-oriented apparatus as our general problem solving architecture. This brings about economies of scale in developing the overall environment.

Constraining the above work to the medical domain results in a narrowing of scope that enhances feasibility. The types of artificial intelligent agents in a medicine-oriented environment include attending physicians, consulting physicians, lab technicians and patients. The world model in this application consists of three components:

1. the model of the human body and its associated physiological and pathological processes (the physical system);
2. the model of clinical practices for each type of medical professional involved (this model specifies the typical professional goals and plans of medical professionals);
3. the cognitive model of the patient, to be coupled with the model of the physical system.

Applications of the medicine-oriented environment may include:

1. Assisting medical professionals in finding needed references, answers to questions, etc. about diseases, their diagnostics and treatment from online sources.
2. Compiling a database of semantically analyzed patient instances drawn from available databases of patient records. These patient instances can be used as training materials, reference materials, or to validate the models being developed in the environment.
3. Compiling a database of clinical practices that can be queried by healthcare professionals.
4. Creating patient information materials, including simulations of what their disease might look like if they continue their current lifestyle, information on how to improve their health given a certain disease, and other patient maintenance-related issues.
5. Providing automatic interactive environments for the use of medical examination boards.

MVP is the first application of the above environment that we have undertaken.

Discussion

How does our modeling differ from other types of modeling?

Computer models can be classified according to several parameters, including whether they are:

A. Stochastic or deterministic

Deterministic Models take no account of random variation and therefore give a fixed and precisely reproducible result. They can be implemented by numerical analysis or computer simulation. Deterministic models are often described by sets of differential equations.

Stochastic Models are mathematical models, which take into consideration the presence of some randomness in one or more of its parameters or variables. The predictions of the model therefore do not give a single point estimate but a probability distribution of possible estimates.

The essential difference between a stochastic and deterministic model is that in a stochastic model different outcomes can result from the same initial conditions.

B.Steady-state or dynamic

Steady-state models use equations defining the relationships between elements of the modeled system and attempt to find a state in which the system is in equilibrium. Such models are often used in simulating physical systems, as a simpler modeling case before dynamic simulation is attempted.

Dynamic simulations model changes in a system in response to (usually changing) input signals.

C. Continuous or discrete

In **discrete models**, the state variables change only at a countable number of time points that mark execution of events and, thus, changes of state.

In **continuous models**, state variables change in a continuous way, and therefore there is an infinite number of states.

Examples of kinds of models include the following:

- A **continuous dynamic simulation** performs numerical solution of differential equations. Periodically, the simulation program solves all the equations, and uses the numbers to change the state and output of the simulation. Applications include flight simulators, racing-car games, and simulations of electrical circuits.
- A **discrete event simulation** manages events in time. Most computer simulations are of this type. In this type of simulation, the simulator maintains a queue of events sorted by the simulated time they should occur. The simulator reads the queue and triggers new events as each event is processed. It is not important to execute the simulation in real time. It's often more important to be able to access the data produced by the simulation, to discover logic defects in the design, or the sequence of events.

- **Agent-based simulation** is a special type of discrete simulation, which does not rely on a model with an underlying equation but can nonetheless be represented formally. In agent-based simulation, the individual entities (such as molecules, cells, trees or consumers) in the model are represented directly (rather than by their density or concentration) and possess an internal *state* and set of behaviors or *rules* which determine how the agent's state is updated from one time-step to the next.

Our approach to modeling is **agent-based, dynamic** and **discrete**. We believe that this approach is the best fit for creating an artificial world in which the agents feature a complex combination of capabilities – simulation of a physical and physiological organism, goal- and plan-based cognitive reasoning, communication between an artificial agent and a human and communication among artificial agents.

Is our model data-driven?

At this time, our model is primarily knowledge-driven, in the sense that it relies on symbolic modeling of causal chains of physiological, pathological and clinical events. Our approach proceeds from the assumption that the experts have created in their minds at least partial models of the above types of events, so that the scientific task is twofold:

1. eliciting from the experts the models that they have and formulate them explicitly in a descriptive theoretical statement;
2. validating (or, if you will, falsifying) these models empirically.

In order to facilitate the second task, one must

1. develop a formal computer simulation of the physiological, pathological and clinical events at the core of both the expert model and the formal theory that reflects it;

2. run the simulation on a representative sample of disease manifestations and either

   3a. have physicians judge whether the simulation is realistic or

   3b. compare the results of the simulation with actual medical records of instances of the diseases in question, their progression and treatment outcomes.

We plan to use both methods of validation/falsification. We have already started eliciting physicians' judgments about the system and plan to start using medical records as soon as possible. Using the medical records for validation is a classical option for making the model data-driven.

Are we using stochastic methods in our modeling?

Stochastic methods are used in our modeling but not as the only or even primary means of modeling. In our initial implementation, due to the nature of the application in which we decided to embody our model, it was essential to avoid randomness in the behavior of the artificial intelligent agent (the virtual patient) for pedagogical purposes – the teacher created patients that behaved in ways that he or she thought are the most relevant and efficient in the educational process. Note that different teachers were able to create different sets of sample virtual patients. Another reason for trying to eliminate randomness was to attempt to provide an even level field for examinees.

With all that, randomness and, therefore, the use of stochastic approaches, is present in the model in several manifestations. First of all, wherever clinical "bridges" were used in describing physiological and pathological processes whose mechanisms are not yet known to science, the clinical expertise of the expert physicians was elicited, formalized and encoded in the model. This expertise is formalized in the model largely through the use of value ranges (instead of single values) of a variety of model parameters (features). Selecting a value from such ranges for a particular instance of a virtual patient was the core operation during authoring virtual patient instances in our initial pedagogical application. To date, the elicitation process has not included judgments of distribution of values within the posited range. But for an application in which the authoring of patients is carried out automatically this stochastic capability can, and will, be added. One way of deriving knowledge for such distribution is stochastic and based on a large-scale analysis of medical records. Such a study is in its planning stages. Also note that expanding the team of expert physicians taking part in this process will inevitably lead to enhancing variability in the descriptions of the physiological and pathological properties, which will introduce even more potential randomness in the model.

Stochastic modeling is also used in several of the modules of the natural language processing component of our model, in cases where encoded knowledge fails to produce complete results.

Is our modeling approach agent-, team- or population-based?

In the classical use of these terms, population-based modeling is typically carried out stochastically and purely empirically. Agent-based modeling is often defined as concentrating on modeling individuals and can be done purely empirically or based on internalized knowledge. In practice, most agent-based approaches model agents that are rather simple and carry out a limited set of actions (methods). For example, a thermometer may be modeled as an agent capable only of measuring temperature.

Our approach does not reject this interpretation of the agent metaphor out of hand. However, we are centrally interested in building causal models of complex whose capabilities in some sense approach or model human capabilities. That is why our model can be called two-level agent-based model: it includes both low-level (e.g., thermometer-

like) and high-level (decision-making) agents. In addition to the above, our interest extends beyond individual agents and to societies (teams) of agents. This is why our high-level agents must be endowed not only with decision-making capabilities vis-à-vis the world (including their own bodies) but also communication capabilities and capabilities of making decisions taking into account the fact that they are members of an agent society. Finally, as the agent societies we are interested in involve both artificial and human agents, it is essential to endow our artificial agents with the capability of communicating in natural language. This latter capability includes both understanding and generating text.

Unlike the majority of recent research into agent networks, our concern is not in allocating resources to collaborating uniform-skill agents who are working toward a common goal, our concern is to manage the interactions of "specialist" agents who must establish their own plans and goals and interpret those of other network agents.

Why do we believe that our work is feasible?

First, our development efforts are targeted toward **specific applications**: there is no attempt to develop a fully generalized, plug-in ready cognitive architecture (like TRAINS/TRIPS), or to implement a broad-coverage, domain-independent dialogue system, or to equip system agents with all of the plans and goals of human beings, or to endow them with the full spectrum of possible character traits (as is done in theoretical approaches to affective modeling), or to model diseases at a grain size any finer than that needed to support the given application. Instead, theoretical and practical advancements are geared toward the near- and long-term future of the specific systems, with infrastructure decisions being made with a long-term view but knowledge support targeted at near-term goals.

Second, the **integrated approach to knowledge modeling** in MVP permits the same ontological substrate to be used for knowledge-based simulation, planning, and NLP, meaning that once knowledge is encoded it is available to all system agents and processors. The OntoSem ontology used in MVP already includes over 8000 concepts, described by an average of 16 properties each; around 7000 of those are from the general domain, with the remaining 1000 devoted to medicine. Moreover, since scripts describing complex physiological and cognitive events are formally part of the ontology, the same scripting language used for physiological simulation (which is already understood by our simulator engine) can be used for planning and dialogue.

Third, the dialogue processing model is grounded in the **OntoSem deep semantic natural language processing system,** which has been under development for over 20 years.

Fourth, the past decade has produced a valuable **body of research** on cognitive engineering, agent networks, planning, plan- and goal-centered dialogue systems, etc. This large body of work includes inventories of needs for intelligent systems, sample

architectures, descriptions of problems encountered, bridges between descriptive, theoretical and implementational work, and reports from the field that provide a good understanding of the current state of the art. In short, this body of work is permitting us to quickly reap the benefits of hard-won insights.

How can we validate our models?

There are three main methods of evaluation and validation of our approach:

- **Validation through use of an application system**. In the case of MVP, experts will manage virtual patients over time and evaluate whether patient responses and outcomes are in keeping with what the experts encounter in clinical practice.
- **Validation using an omniscient view of the system**. In the case of MVP, experts will manage patients while having access to the physiological properties that are used by the simulation but are unavailable to the user in a typical training scenario. The experts will evaluate whether the model encoded in the system is compatible with their own mental models.
- **Validation through comparison of real patient records with our models**. We could semantically analyze patient records, formulate their information in terms of our disease and treatment models, and see if our models cover the actual disease manifestations and physician decisions regarding when and how to treat.

How can our theory be falsified?

A theory is scientific only when it is falsifiable. We believe that our theory of modeling virtual patients is indeed scientific. Our theory is two-fold. One facet of the theory is the actual physiological and clinical knowledge encoded in the system. The other facet of the theory consists of the cognitive models of the physiological and reasoning agents.

The former facet of the theory can be falsified if expert physicians determine that a) the progression of diseases in the virtual patient does not correspond to their experience with such diseases; b) the reaction of the virtual patient to treatments does not correspond to their experience; or c) the treatments and diagnostics selected for a particular case or disease are not deemed appropriate. The cognitive model can be falsified if the behavior of the system agents is judged incoherent or unexplainable. (This behavior includes natural language dialog capabilities.)

Method and Application Comparisons

How does our work differ from data mining?

Data mining is primarily data-driven, with relatively little encoded knowledge brought to bear on the process of extracting useful “nuggets” of information from very large information streams. Often data mining concentrates on what is known as metadata, data about information (e.g., the author of a text, the date of its publication, the source where it was published, etc.). More advanced applications of data mining involve some analysis

of the data extracted, described as "knowledge discovery." Much of data mining uses databases as input. The core methods of data mining are empirical and statistics-oriented, though they can use prerecorded knowledge. We expect to use a variety of data mining techniques in the future, when we have collected a sufficient substrate of static knowledge and text processing capabilities to allow semantics-enhanced data mining over open text.

How does our work differ from expert systems?

Many of the AI systems in medicine have been expert systems, defined as systems that stand in for an expert, most typically by offering diagnostic assistance (for a list of over 50 such systems, see http://www.computer.privateweb.at/judith/index.html). Expert systems are, as a rule, not grounded in simulation and do not provide for extensive interaction. MVP is not an expert system, as it does not stand in for a physician. Furthermore, MVP does not follow the conceptual or architectural lead of classic AI systems, which were typically large and difficult to maintain sets of production rules. Instead, MVP has a proportionally far smaller inventory of rules that are more structured and aggregated, and contain control information that boosts system efficiency.

How does our agent society differ from most agent organizations?

A large number of different types of organization are in use in the area of multi-agent systems. Horling and Lesser (2005) distinguish hierarchies, holarchies, coalitions, teams, congregations, societies, federations, markets, matrix organizations and compound organizations. Our approach uses some features of agent team and some of agent society. Horling and Lesser define an agent team as follows: "An agent *team* consists of a number of cooperative agents which have agreed to work together toward a common goal. In comparison to coalitions, teams attempt to maximize the utility of the team (goal) itself, rather than that of the individual members. Agents are expected to coordinate in some fashion such that their individual actions are consistent with and supportive of the team's goal. Within a team, the type and pattern of interactions can be quite arbitrary but in general each agent will take on one or more roles needed to address the subtasks required by the team goal. Those roles may change over time in response to planned or unplanned events, while the high-level goal itself usually remains relatively consistent." The above, of course, is applicable, under some interpretation, to most cooperative multi-agent systems. Agent societies are understood as a looser organization, with a "characteristic set of constraints they impose on the behavior of the agents, commonly known as *social laws*, *norms* or *conventions*. These are rules or guidelines by which agents must act, which provides a level of consistency of behavior and interface intended to facilitate coexistence."

There is, however, a big difference between our environment and the environments described in Horling and Lesser: the environments of the kind we build, by definition, involves a human as one of the agents.

How do our models differ from those based on canned scenarios?

Most currently available clinical decision-making systems are not grounded in simulations. Instead, they provide trainees with the opportunity to work through decision trees that target key decision points in the process of diagnosing and treating a disease. MVP, by contrast, offers trainees a more open-ended choice space and a richer scope of interactions, which more closely parallel the demands of actual medical practice.

What are our differences with other projects/systems?

CIRCSIM

CIRCSIM-Tutor is a system whose focus has fundamentally shifted since its inception (Evens and Michael 2006). Initially, under the name MacMan, the system was a mathematical model of the baroreceptor reflex that could be explored by students but provided no feedback. However, it was found that the lack of feedback made it a non-optimal teaching tool, which led to the development of the first spin-off system, Heartsim, which offered limited feedback. With the Heartsim system, developers realized that the mathematical model was not being exploited and that the most effective teaching was based on stored correct predictions rather than real-time calculations using the mathematical model. As such, the final system, CIRCSIM-Tutor (which is still under development), does not actively use the mathematical model: the dynamic aspect of the system is constrained to the tutoring process itself. To summarize, this system went from offering students a dynamic mathematical model with no tutoring support to offering them tutoring without the dynamic mathematical model. In the MVP system, by contrast, the autonomous functioning of MVPs is, and will remain, no less important than the mentoring aspect of the application.

The developers of CIRCIM are pessimistic about the prospects of automatic tutoring in a less than highly constrained realm:

> "When we started the CIRCSIM-Tutor project 15 years ago, some experts in the field argued that student modeling was too difficult to be worth the trouble; some even classified the problem as totally intractable… Anyone who observes human tutors in action, on the other hand, must recognize that they base decisions at all levels, from the choice of the next problem to present to the student to what kind of hint to provide, on their model of the student… Joel Michael and Allen Rovick were so convinced of the crucial importance of modeling that they picked the CIRCSIM domain [the baroreceptor reflex] for our tutor largely because they felt that it would be easy to construct a good student model in this subject area … They are… convinced that it is important to build a comprehensive model before starting to tutor, to ensure that the tutor can begin by attacking the most important of the student's conceptual difficulties" (p. Evens and Michel 2006: 252-3)

Undoubtedly, selecting a narrow purview facilitates domain modeling, student modeling and the automation of tutoring support; and, all other things being equal, one would expect better near-term results from a more highly constrained system. However, all other things never really are equal: there is a real-world need for simulation and tutoring in the broad domain of diagnosis and treatment of disease, and it is this need that has set the agenda for our research and development. While task-driven projects necessarily involve unknowns, they also promise exciting new horizons both within the targeted application and beyond it.

TRAINS/TRIPS

TRAINS and its successor TRIPS (hereafter, TT) are projects devoted to the study of natural language dialogue between an intelligent agent and a human as they collaborate on a planning task – specifically, cargo transport. The interconnection between dialogue and planning is key in these projects – as it is in MVP. Of particular interest for us is the work of David Traum, who has created a bridge from theoretical linguistic descriptions of dialogue to practical implementations of dialogue systems within a plan-oriented cognitive architecture. We are not, however, attempting to implement a system just like TT. In fact, there are at least four differences between the two environments: 1) TT is based on a generalized architecture that must manage compatibility issues between plugged in components, whereas the architecture of MVP is system specific, thus reducing one type of complexity; 2) TT does not pursue the depth of semantic text processing that MVP does; 3) TT does not incorporate a sophisticated simulation system; 4) TT does not include a tutor (the domain does not call for one). By having a broader scope of work devoted to a more complex domain, MVP clearly involves more risk, but also promises a bigger payoff: a training and mentoring system to support the next generation of physicians. For further information, see the project website: http://www.cs.rochester.edu/research/trains/

Knowledge Machine

KM is a knowledge representation and reasoning system developed by a group at the University of Texas. It permits users to encode knowledge, query the database about the knowledge, and run simulations that would indicate what the state of objects in the given world would be if certain events took place.

KM and OntoSem are both frame-based knowledge representation languages, the former focusing on logic-based reasoning and the latter on text processing and, as of late, simulation. Some points of similarity among the environments are:

- a three-tiered distinction between concepts (called 'classes' in KM), instances, and a hybrid entity that lies somewhere in between (in OntoSem this hybrid is called an ontological instance; in KM it is called a proto-instance). While the definitions and distribution of these three types of entities is not precisely parallel

in the two environments, the theoretical and methodological reasons for making such distinctions overlap

- the use of primarily first-order logic in describing entities (concepts, instances and those hybrid entities) using properties and fillers
- the possibility of placing complex fillers in slots, nesting such descriptions to any depth, and tracking coreferences within such nested structures (permitting the construction of scripts and prototypes)
- the use of a large inventory of properties, and the possibility of adding more as an acquirer deems necessary
- the support of knowledge expressed in if-then statements
- the support of reasoning for sophisticated applications.

The main difference between KM and OntoSem involves the focus of development effort and the applications supported. KM is a freely distributed environment that permits users to created databases and reason over them using theorems of predicate logic. OntoSem, by contrast, focuses on natural language processing in general as well as simulation in the medical domain.

Other Knowledge-Based Simulation Work in Medicine

Recent literature includes two projects that offer interesting points of comparison with the MVP system. Walton Sumner and his associates (Sumner and Hagen 2006, Sumner et al. 1996, 1996, Marek et al. 1996) developed a system for the purpose of advancing medical certification procedures of the American Board of Internal Medicine beyond the level of multiple-choice questions. The system relies on simulating a patient with possibly multiple co-existing conditions and allowing the examinee to intervene. The system addresses the issue of automatically creating instances of virtual patients by starting with a selected disease (or diseases) and probabilistically creating a medical history for each patient instance based on knowledge about the incidence of this disease (or diseases) in the population. This emphasis on generating differentiated patient histories is due to the perceived need for providing a secure testing environment in which patient instances are not reused. While the knowledge in the system covers “health states” of a virtual patient and causal and temporal connections among them (with their associated properties and symptoms), this knowledge does not yet support a realistic simulation of a virtual patient’s physiological processes. The probabilistic nature of much of the operation of this system suggests the use of Bayesian networks as the underlying representational mechanism, which adds complexity to both knowledge acquisition and processing.

Of the many differences between this system and the MVP system, we believe the most salient are: a) the Sumner system focuses only on evaluation, whereas MVP centrally includes training; b) MVP provides for following patients over time without the need for any explicit intervention; c) MVP achieves a level of realism in patient simulation not attempted by the Sumner system, particularly by virtue of endowing the VPs with a cognitive agent capable of perception (sensory and language), reasoning (goal-oriented decision making) and action (physical and verbal); d) MVP permits the user to participate

in an agent network that involves other human-like agents, including a dedicated tutoring agent; e) MVP uses stochastic methods much less than the Sumner system in order to keep all data and processing explicitly controlled and inspectable by developers.

Amigoni et al. (2003) describe a multiagent system for the modeling and regulation of physiological phenomena, specifically, for regulating the insulin and glucose levels in diabetes patients. This system relies on what the authors call the "anthropic agency" architecture, "a powerful paradigm to develop control systems for physiological processes shaped as multiagent systems." (p. 310). The architecture consists of the traditional steps of perception (called "knowledge extraction", implemented using a team of identical "extractor agents"), reasoning ("decision making", implemented using a team of identical "decisional agents") and action ("plan generation" with their team of "actuator agents"). Communication among agents from different groups is mediated through messages on blackboards. Communication among agents of the same group is prohibited. The blackboards are serviced by their own agents, and the overall architecture also includes a database agent and "majordomo" agent responsible for communication with "both the technical expert, who can modify the composition of the system by adding and removing agents, and the medical expert, who can inspect and tune the parameters and the functioning of the system." (p. 313). The actual implementation involves one extractor, one actuator and two decisional agents and essentially simulates the fluctuation of just two properties – glucose and insulin levels with two outside influences – food intake and physical activity.

All the agents in this system are what we call low-level agents. The physiological model has a narrowly directed coverage, no cognitive abilities are simulated for the virtual patient and no network of high-level agents simulating human capabilities is introduced. In general, the complexity of the domain knowledge is not the main focus of this work. The main thrust of the paper is the discussion of the algorithms that embody the agents and the algorithms supporting their communication (the negotiation mechanism). This is understandable considering that the intended application of the anthropic agency is intelligent prosthetics – implanting an intelligent insulin supply regulator in diabetic patients.

Our work, by contrast, is devoted to immersing human users in simulated environments. Therefore, we concentrate on the breadth of coverage and realism of the simulation and postpone the discussion of such architectural and control issues as resource allocation, efficiency of agent collaboration and task scheduling. This deferral is made possible by making a variety of simplifying assumptions in our system. For example, we assume that the scheduler, executor and the demons have all the actual time they need to complete all their operations at a given agenda time slot. In other words, virtual time is not equated with real time. Our simulations at this time do not suffer because of this assumption because we have been dealing with chronic diseases that unfold over long periods of time. We will revisit this simplifying assumption when we will start modeling acute diseases where durations of crucial events can be very short. Similarly, our application environment allows the simplifying assumption of full availability of not only computational resources but also resources in the world being modeled. By the same

token, our system's quality and utility will not suffer if we do not concentrate on scheduling particular specialists or lab technicians for particular tasks. Since the one agent in our system that is intended to operate under uncertainty is the human user, we can avoid the complexity of introducing probabilistic reasoning engines.

Should the need for removing the simplifying assumptions arise in our future work, we intend to use advances in multiagent coordination, which is a very active area of research (cf., e.g., Lesser et al. 2004 or Pynadath and Tambe 2003). Our work is complementary to such studies. Using multiagent coordination results, we can improve the efficiency of our systems, while the coordination testbeds will benefit from being able to test the coordination algorithms and architectures using the knowledge-rich and heterogeneous agents developed in the MVP system.

Other System Comparisons

Several existing systems and projects share a small number of characteristics with our approach. In what follows, we give a very brief survey of such systems.

Realistic Simulation. One type of computer-aided training involves technical task trainers. These focus on training a specific technical step, with little or no cognitive simulation. Like our environment, they aim to be sophisticated and lifelike. For example, manikins to teach the care of infants and adults have been developed by Laerdal, Inc. ("SimBaby", http://www.laerdal.com/; "SimBaby") and Meti, Inc. ("The Human Patient Simulator", http://www.meti.com/), respectively.

Cognitive Training. Among the computer methods to train decision-making skills are systems based on decision trees that embody diagnostic and treatment algorithms at the case level. These include no biomechanistic processes, and the user is limited to selecting one of the pre-scripted options at fixed points in case. MedCases, Inc. (http://www.medcases.com) is an e-learning company that develops patient interaction scenarios for continuing medical education. Although these are far more "canned" that the ones in MVP, they too seek to train cognitive capabilities.

Hybrid Models. A well-known simulation project is the Virtual Soldier (http://www.virtualsoldier.net/), which simulates the human thorax in the context of penetrating trauma. It combines the Foundational Model of Anatomy with a stochastic physiological knowledge at the cell, tissue, body and population levels. Although Virtual Soldier differs from our work in several ways – as by focusing on the short-term treatment of trauma rather than the long-term treatment of patients with ever-changing disease states – the approach integrates various types of knowledge, as does ours.

Intelligent Behavior in Changing Circumstances. Like the well-known expert systems (e.g., Mycin Shortliffe XXX), MVP shows intelligent behavior in ever changing circumstances. However, unlike traditional expert systems, our system is grounded in simulation, stresses language-based interaction, uses the same knowledge bases for both

simulation and interaction, and permits free-form interventions – all of which involve innovative uses of adaptive computing.

Large Population of Patients. This is a feature which is shared by our approach with that of Sumner and Hagen (2006).

Medical Tutoring. The CIRCSIM project (Evens and Michael 2006) concentrates on tutoring in a medical domain and involves natural language dialog – just like MVP. However, CIRCSIM-Tutor currently does not incorporate simulation, and it covers only one specific medical condition, the baroreceptor reflex – the body's rapid response system for dealing with changes in blood pressure.

Knowledge-Based Dialog. Susan McRoy at the University of Wisconsin, Milwaukee has been developing a dialog system in the framework of a tutoring environment for medical students (e.g., McRoy et al. 1997). She shares our belief in the need for knowledge for language processing but focuses on dialogue issues without detailed specification of physiological and pathological states.

Surgical Innovation
Volume 15 Number 1
March 2008 7-16

10.1177/1553350608316573
http://sri.sagepub.com
hosted at
http://online.sagepub.com

Novel, Web-Based, Information-Exploration Approach for Improving Operating Room Logistics and System Processes

Paul G. Nagy, PhD, Ramon Konewko, MS, Max Warnock, Wendy Bernstein, MD, Jacob Seagull, PhD, Yan Xiao, PhD, Ivan George, BS, Adrian Park, MD

Routine clinical information systems now have the ability to gather large amounts of data that surgical managers can access to create a seamless and proactive approach to streamlining operations and minimizing delays. The challenge lies in aggregating and displaying these data in an easily accessible format that provides useful, timely information on current operations. A Web-based, graphical dashboard is described in this study, which can be used to interpret clinical operational data, allow managers to see trends in data, and help identify inefficiencies that were not apparent with more traditional, paper-based approaches. The dashboard provides a visual decision support tool that assists managers in pinpointing areas for continuous quality improvement. The limitations of paper-based techniques, the development of the automated display system, and key performance indicators in analyzing aggregate delays, time, specialties, and teamwork are reviewed. Strengths, weaknesses, opportunities, and threats associated with implementing such a program in the perioperative environment are summarized.

Keywords: graphical dashboarding; business intelligence; scorecarding; information visualization; quality

Management of the modern perioperative environment is a challenging act of balance and orchestration that often tilts perilously close to chaos. Many unforeseen delays (including, but by no means limited to, patient transport; case cart preparation; consent forms; and slow turnover) can trigger a cascade of events that escalate throughout the day, resulting in frustration for physicians, staff, and patients. The cumulative effects of many small and interacting delays keep the operating room (OR) from running at peak efficiency and can, in some cases, contribute to more serious errors in management and care.

From the Departments of Diagnostic Imaging (PGN, MW), Anesthesia (WB, YX), Surgery (JS, IG, AP), University of Maryland School of Medicine; and Department of Surgery (RK), University of Maryland Medical Center, Baltimore, Maryland.

Address correspondence to: Paul G. Nagy, PhD, Department of Diagnostic Imaging, University of Maryland School of Medicine, 22 S Greene St, Baltimore, MD 21201; e-mail: pnagy@umm.edu.

Managers trying to address these delays in an ad hoc fashion find themselves playing “whack-a-mole” with serial problems; as soon as one glitch is resolved, another rears its head. The result is a reactive approach that focuses on immediate problems at the expense of the time and effort needed to identify root causes and long-term solutions that will prevent recurrences. The good news is that various routine clinical information systems now gather enormous volumes of data that surgical managers can access to create a seamless and proactive approach to streamlining operations and minimizing delays.

However, the process of leveraging these data in support of routine improvements presents its own challenges, particularly in rethinking traditional reporting and analysis techniques. In the past, paper spreadsheet–based reporting methodologies have been used in management meetings, an approach that is insufficient to handle or analyze even the broadest trends in the increasingly large volumes of

useful data collected. This traditional, tactical approach most often focuses on only the short-term history of operations and fails to identify the small, recurrent delays that may occur across services.

Transparency and a broad scope of accountability are widely recognized as hallmarks of high reliability and dedication to quality in health care organizations.[1] These values, along with an emphasis on verifiable metrics and automated means of collection and assessment, have figured in significant advances in operations research in the management of the surgical environment in the past 5 years.[2-6] These advances contribute to the fulfillment of important goals of surgical units, including patient safety, access to ORs, economic efficiency, waiting time, and staff satisfaction.[6] Moreover, they have provided novel information about what factors contribute to which specific quality goals. Simulation studies using operations research, for example, have indicated that although immediate quality improvements in patient safety, waiting time, and satisfaction on the day of surgery should be a primary focus, only longer term decisions on staffing will provide economic efficiencies.[6,7] Thus, in general, reduction in turnover time will not result in increased volume,[8] but access to historical data and application of operations research methods can point to staffing solutions that will optimize economic efficiency.[9]

Our goal, as part of a grant on the OR of the Future from the Telemedicine and Advanced Technology Research Center, was to accelerate the adoption of these advances by providing an automated, holistic view of operations that would enable the managers to discover patterns and causes of delays. We created a Web-based, graphical dashboard that could be used to interpret clinical operational data, could allow managers to see trends in data, and could help identify inefficiencies that were not apparent with more traditional approaches. This dashboard was designed to provide a visual decision support tool that would also assist managers in pinpointing problem areas in which the greatest benefits could be achieved by applying time and energy toward continuous quality improvement.

What is Business Intelligence?

The field of business intelligence, sometimes referred to as business analytics, is the utilization of data warehousing, data mining, modeling, and forecasting to aid in managerial decision support systems.[10,11] Business intelligence is defined as extracting useful information from the data generated by operational systems of an enterprise.[12] Many top corporate executives use business intelligence–generated electronic scorecarding and dashboarding methodologies to manage their operations with real time–decision making support. A 2007 Gartner Inc worldwide survey of 1400 chief information officers ranked business intelligence as the number 1 technology priority for remaining strategically competitive.[13] Business intelligence methodologies extend directly to consumers in certain markets. Financial Web sites provide individual investors with extensive research and graphical performance analysis of publicly traded companies.

Large academic medical centers, which often generate revenues in excess of US$200 million, are in the same financial league as the medium-to-large businesses in which dashboards are commonplace. Yet few medical centers have invested in the development and routine implementation of tools to analyze and improve the efficiency and effectiveness of perioperative management and operations. Many ORs continue to “fly blind” regarding concepts such as indexing and performance measurement.

Internal graphical dashboards have been proven in other environments to provide useful and productive platforms for continuous quality improvement. The Six Sigma quality methodology, for example, frequently uses dashboards for process management.[14] A dashboard can provide a consistent framework of defined metrics, known as key performance indicators, that aid in defining and redefining quality and goals, as well as offering quantifiable data on achievements.[15] Evidence of consistent improvements through a public dashboard is then used to help align the various parts of an organization to target enhanced performance.

The Potential of Information Visualization

Our project was designed to provide a visual knowledge exploration system to assist managers and senior leadership in understanding trends and patterns. The tools used in this system provide interactive views of data at various granularities and in a series of graphical or tabular formats. These tools can quickly and with minimal user effort impose various

types of analyses on the full data set or on interactively selected subsets of data.

The goal of a visual knowledge exploration system is to provide tools that facilitate interaction with information in an easy, transparent, and meaningful manner. A well-designed graph can tap into the pattern-recognition capabilities of the human visual system. In certain types of patterns, human vision can identify a unique (outlier) value within 200 milliseconds, regardless of whether few or many data points are present.[16,17] However, this ability is entirely dependent on the manner in which the pattern is displayed. Proper visual display is crucial to the use of large data sets for complex decision-making support. The optimal type of data display has been the focus of a substantial body of literature and reporting, and the definitive answer changes as rapidly as new technologies enter the information arena.[18-22] Some studies suggest the superiority of graphical formats (bar charts, pie charts, etc) over tabular presentation (data tables) for certain tasks, whereas the reverse is true for other tasks. More recent work indicates that a constellation of factors must be considered in determining the most advantageous data set display formats, including type of task, underlying structure of the data, and the knowledge level of the users.[23]

What Is the Problem with Paper-Based Reporting?

The benefits of graphical dashboarding can be appreciated more fully by looking at the limitations of traditional paper-based reporting in identifying and managing ongoing operations challenges. Understanding these limitations is important because in many institutions paper-based reporting is so engrained into routine practice that clinicians and perioperative managers may find it difficult to take the steps needed to adapt to other methodologies. Paper-based reporting management systems are limited in the following areas:

Time

Significant time is required to gather information from various sources and compile reports by hand. Decision making is a time-sensitive activity that requires actionable information. Decision making, a process that that should be based on fresh data, is adversely affected when time simply does not permit the preparation of all possible permutations of analyses that might be informative and useful.

Effort

The inherent limitations of paper restrict the number of questions than can be asked and tend to generalize rather than to drill down in areas of analyses. Expanding the scope of a paper report requires extra labor. Most often, the result is a trade-off between the time required to generate the report and the quality of effort required in preparing the results for analysis.

Hindsight

One of the most frustrating characteristics of paper-based reports is that they provide only answers to questions that were identified before the management meeting and discussion. New questions asked during the meeting must be tabled until the next meeting so that analysts can gather the new information required. These tabled questions prevent more purposeful discussions about the data and leave managers with limited information to support decision making in the short term.

Scope of Report

The amount of information that can be contained in a paper-based report is limited, as is the amount of information that can be reviewed within a reasonable amount of time. Selectivity becomes a necessity; yet, it is difficult to predict which questions managers will have during any given meeting. Attempts to broaden the scope of paper-based reports can be both time consuming and problematic; the larger the amount of data in a paper report, the more difficult it is to find any specific piece of information.

Granularity

Aggregate statistics do not allow the user to drill down to understand the underlying distributions to evaluate credibility. Mean statistics offered in most paper-based reports are unreliable when describing nonnormal distributions of data. A single chart or table on paper can show only one view, and it is difficult to present both overviews and detailed information in a single presentation. Showing trends can

obscure source data, where showing only source data can obscure trends. Including both or all in a paper-based report can be labor/time intensive to review and is impractical as a routine practice.

Multiple Versions of Truth

Different groups generating separate reports or even the same reports at different times can result in conflicting operation directives that can add to confusion in effective decision making. The human intervention inherent in paper-based reporting can also introduce bias that may lead to data analysis errors. Moreover, the passage of time between the collection or analysis of data and final reporting may mean that no direct links exist between current operational data and the paper-based report.

Dashboard Design

Organization of the Web Site

The Web site was designed to allow analysis from several perspectives. One of the principle mantras of information visualization and data discovery, identified by Shneiderman,[24] is the ability to overview first, zoom, filter, then details-on-demand. The ability to view data from multiple perspectives assists and increases confidence in decision making. The user accesses each category via a navigation bar of tabs at the top of the Web site. As the user navigates through the system, a trail (called a breadcrumb; Figure 1) is displayed to illustrate how the user navigated to that point and to allow easy backtracking.

To create a concise visualization environment, we created graphs that were clickable within a drill-down interface to provide fast and intuitive zooming and filtering of data. These graphs also provide detailed information about data points when the user hovers over a data marker with the mouse. One of the core requirements for a management dashboard is that it be Web based to allow secure access to all authorized users from any location at any time.

Standard data warehousing techniques were used extraction, transformation, and loading. The database was populated by parsing a text file in a comma-separated delimiter format provided by Cerner SurgiNet Surgical Information System (Kansas City, Missouri). The text file was converted to ANSI Structured Query Language commands as

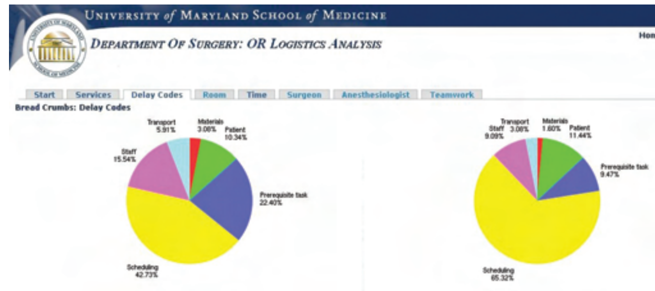


Figure 1. A, Pie chart analysis of number of occurrences of each delay type. B, Pie chart of the relative amount of delay caused by each delay type.

inserts using the MySQL database administration utility PhpMyAdmin. Data were anonymized for patient information because the focus of the system was operational efficiency, not identifying specific patient-associated incidents. In all, 6 months of operational data were uploaded into the system, incorporating performance statistics on 7807 cases on 8 MB of disk space on the server. These cases incorporated all of the operating rooms with both inpatient and outpatient admissions.

Identifying Key Performance Indicators

The American Association of Clinical Directors derived a common glossary of the exact meaning of times used for scheduling and monitoring surgical procedures.[25] Time stamps were extracted from the clinical database for scheduled start time, time at which the patient enters the operating room, the time at which surgery begins, surgery end time, time at which the patient leaves the room, and the turnover time of the room. As a hospital policy, when a case begins 15 or more minutes later than scheduled, the circulation nurse must specify a reason for the delay. These performance data are combined with data from the case such as the room, surgeon, anesthesiologist, case number of the day, and the service.

A total of 43 delay types were identified as reasons for delays. These were grouped into general root causes of materials, patient, prerequisite task(s), scheduling, staff, and transport. At the top of the delay analysis page, as shown in Figure 1, pie charts demonstrate the relative number of delays per root cause and their cumulative impact in time. Although some delays are not numerous, they might have a

large effect on the operations of an OR. The user clicks on the pie chart, selects a root cause, and presents with an analysis page of all the underlying delay causes for that root cause, broken down in the same way by relative number and impact. By selecting a delay cause, the system moves to show a breakdown by specialty, displaying the number of incidents and their average delay times. Selecting a service displays all the cases, and by selecting a case, the details for that case can be displayed. Within the span of 4 clicks, a user can drill down from all the cases in the database to the details of an individual case. The delay analysis tool is useful in understanding the cumulative cost of systemic delays and which specialties are most affected by them.

Temporal Analysis

The temporal perspective provides a daily tactical review of cases to determine over a specified period of time which ones were delayed and why. To present the utilization levels of the ORs in a given day, we used a polar chart showing cumulative room utilization as a function of the hour of the day (Figure 2). This is useful for look at the relationship between room utilization and staffing levels. The user can drill down to the specifics of a single case or can choose to look at data grouped by room or specialty. System delay types, such as transport issues, can affect multiple rooms and specialties across suites of ORs over different periods of time.

Service analysis focuses on key performance indicators within each specialty. Medical specialties within the OR have widely differing dynamics for case efficiency, utilization, turnover, and case length based on a number of factors, including but not limited to procedure complexity and patient acuity. For some types of data analysis involving services or subspecialties, bubble charts provided a useful way to organize data (Figure 3). The bubble chart plots each service by its average case length in the x axis and average delay duration in the y axis. The size of the bubble for each specialty is directly proportional to the number of cases performed. The more cases a service performs, the larger becomes the diameter of the bubble.

For each specialty analysis, the site provides histograms for case length and delay duration. Histographic analysis is useful in determining the distribution type, the spread of the distribution, and the existence of outliers that may distort statistical analysis.

Another display generated was a scatter graph of all cases plotted by their scheduled case lengths

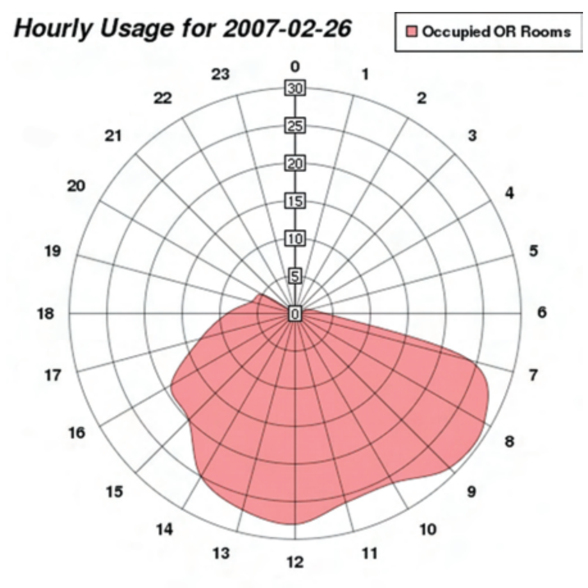


Figure 2. A polar chart of room occupation as a function of the hour of the day. OR indicates operating room.

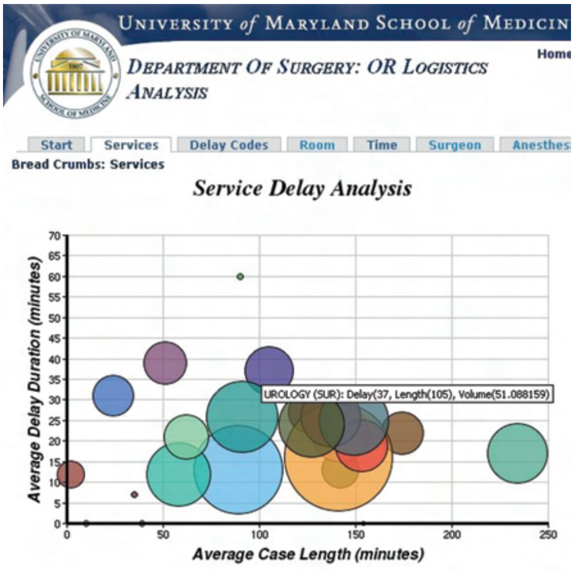


Figure 3. A bubble chart of the services within surgery plotted along their average case length versus their average delay duration. The size of each bubble is proportional to the number of cases performed by the service.

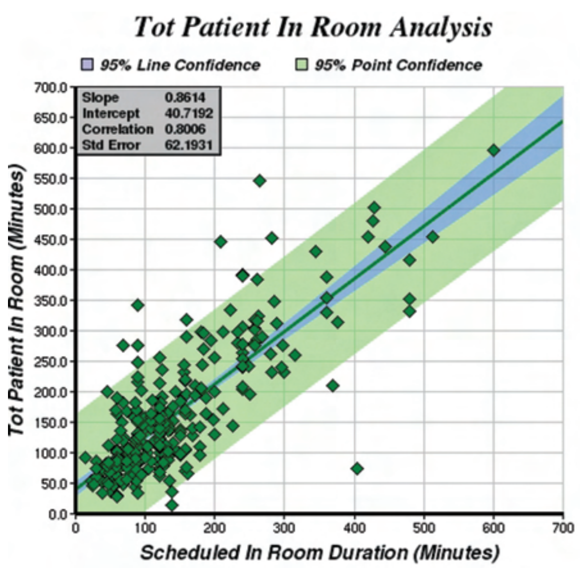


Figure 4. A scatter plot with confidence banding plotting the scheduled in room duration versus the actual patient time in the room.

compared with actual duration (Figure 4). Regression analysis shows potential correlations along with graphical bands illustrating confidence intervals for the line and the points. This index of predictability of scheduling is especially useful in identifying and drilling down on the outlier cases to understand their causes of variance. Each diamond represents an individual case, and by clicking on a diamond, the details of that case are displayed (Figure 5).

Teamwork Analysis

With a surgeon and principal anesthesiologist assigned to each case, we can display delay causes for those cases. As shown in the spider graph in Figure 6, delays are grouped by and aggregated by the blue bars. The farther were the bars, the greater were the number of occurrences for that root cause. The overlapping orange are the average delays by root cause for all physicians in that specialty, normalized by the number of procedures done by that physician.

Using this teamwork analysis, it is possible to identify specific teams that appear to work well together and those that are not routinely time efficient. Other factors, of course, must be considered in reviewing these data, and it would be difficult to determine root causes for efficiency or inefficiency in a specific case. However, this knowledge may provide strategic information that could contribute to what business intelligence experts call a discovery cascade.

Internal Case Number 7694

Or Group	GENERAL OR
Patient Type	Inpatient Admission
Delay Reason	No Delay
Scheduled Start Time	Thu May 03 10:30:00 -0400 2007
Patient In Room Time	Thu May 03 11:07:00 -0400 2007
Patient Out Room Time	Thu May 03 12:22:00 -0400 2007
Surgeon	[redacted]
Speciality	UROLOGY
Case Start Time	Thu May 03 11:26:00 -0400 2007
Or Room	G08
Anest	[redacted]
Sched Anest Type	General Mask
Surgery Start Time	Thu May 03 11:26:00 -0400 2007
Surgery Stop Tme	Thu May 03 12:08:00 -0400 2007
Nurse Unit	W5A Surg
Admit Time	Wed May 02 19:05:00 -0400 2007
Scheduled Or Room	
Anest Type	General Mask
Discharge Time	Fri May 04 12:51:00 -0400 2007
Exp Would Class	Contaminated
Patient Age	[redacted]
Wound Class	Contaminated
Sched Anest	-
Case Level	Elective
Case Minutes	42
Tot Patient In Room	75
Total Patient In Room Anes	0
Total Patient In Room Surgery	19
Total Setup Minutes	35
Total Surgery Minutes	42
Total Delay Duration	0
Turnover Minutes	67
Scheduled Case Duration	405
Case Number	2

Figure 5. A detailed report of a given case. OR indicates operating room.

Results

Results of an initial rollout of the Web system were assessed through interviews with senior management. This included discussions with the chief medical officer, chief operations officer, chief nursing officer, chairs of surgery and anesthesiology, and several perioperative managers. In presenting this potentially disruptive tool to management, we performed a strengths, weaknesses, opportunities, and threats analysis (known in business intelligence parlance as a SWOT analysis) to classify their observations.

Strengths

Our Web-based approach was seen as a powerful tool that would aid management in identifying systemic, process-driven root causes for delays and other problems and that had the potential for positive

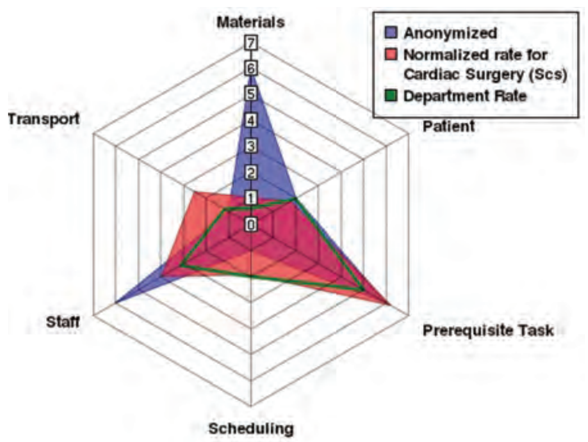


Figure 6. A spider graph of the number of delays broken down by delay type and plotted for the entire department and the section.

effects on the culture of the organization. Among the positive aspects they cited were: (1) this approach turns traditional paper-based data into knowledge and presents this knowledge in easy-to-digest chunks; (2) the dashboard is independent of any single vendor; (3) if used with a data repository, it has the potential ability to link data from different information systems; (4) it provides a systemic view that can calculate the total costs of root causes; (5) it provides a quick visual way to target improvement; (6) additional metrics can be added at the request of management with minimum programming effort; (7) visual displays made outliers and trends more easily identifiable and rendered distributions more easily understood than standard aggregate statistics.

Weaknesses

The reviewers identified 4 areas of weakness and potential improvement for the Web-based system.

Timeliness. Depending on the method of data extraction, data may not be live or near live. If data are provided via an upload, they will be only as recent as the last event. The dashboard optimally should have an Open Database Connectivity connection to a clinical data repository (CDR) or similar copy of the live production environment. The update schedule of the CDR will determine the timeliness of the dashboard.

Personnel resources. Skilled personnel are required to build and maintain a graphical dashboarding application. A surgical informaticist is a good choice as they have the clinical domain experience combined with the principles of management and information technology. This person needs to guide the development of metrics using clinical knowledge to extract meaningful and relevant data. This individual can also play a crucial role in bridging the cultures among health care providers, information technology specialists, and business process managers.[26]

Hardware resources. Hardware resources include access to server space with sufficient processing power and storage to handle a large database. The database storage space required is minimal; however, the central processing unit that drives the data mining must be powerful.

Management training. The introduction of analytics and acceptance of business intelligence practices within a group, particularly one that already has a long-engrained operations process, cannot be accomplished overnight.[11] An investment of time and effort is required and involves education of managers on the use of these tools and the ways in which they can be incorporated into decision making. In the process, the focus of the managers and the entire organization should change from trying to understand the latest event to looking at trends within the data to predict what will happen next and to identify ways to achieve the best possible results.

Opportunities

A dynamic surgical block utilization chart with easily available drilldown, as created in our project, allows surgical chiefs to continually monitor utilization. The drilldown permits them to see which days are being underutilized and by whom and points to immediate courses of action rather than waiting for end-of-year retrospective and analysis. When competition is fierce for OR time, this transparency can be extraordinarily valuable for surgical practices.

Another potential benefit can accrue from matching staffing with caseload to optimize OR efficiency.[27] Cases in overutilized time are 1.75 times more expensive than cases during normal staffing hours, the goal is to match caseload with staff.[28] The dashboard tool pulls in scheduling data from the clinical information system and can display the number of projected cases at 1-hour intervals. The dashboard can also use retrospective data on add-on

cases to estimate a caseload probability by the hour. To maximize efficiency, a user input can be created whereby a manager or charge nurse may enter data on staffing levels, helping to match caseload to staffing.

Along with the scheduling efficiency, OR senior staff and managers may want to match clinical proficiencies with cases. Displays can be created to show circulator/scrub combinations with surgical specialties and case types similar to the surgeon–anesthesia graphs presented. This gives opportunities to the managers to maximize good teams and to identify teams that need improvement.

Many opportunities are available for benchmarking between services and between organizations. The only limitation is the ability to capture data from an information system or network, to apply a meaningful analysis, and to provide an easily understood graphic for the appropriate audience.

Current Procedure Terminology (CPT) codes would be an important additional piece of information because case efficiency should be benchmarked against similar cases. For example, cardiac thoracic cases have long turnover times because of the degree of complexity involved in setup of equipment, drawing of drugs, patient preparation, etc. National benchmarking can be imported to compare against the organization's benchmarks reports for CPT and disease-related groups, morbidity and mortality, length of stay, and complications and can be presented in an easy-to-navigate and easy-to-understand visual.

Opportunities for assessing clinical outcomes include measures such as infection control, preoperative antibiotic compliance, unplanned returns to the OR, staff compliance on chart quality, timeliness, completeness, staff arrival time, etc. Financial reports can include direct costs, indirect costs, contribution margins per case/specialty, labor costs, supply costs per case, and metrics associated with defining and monitoring best practices.

Threats

In all, 2 potential threats to a system, such as the one we devised were identified by the interview group and by our own developers.

The first threat is in the area of data quality and integrity. The data retrieved for our clinical information system have 2 sources of origin. First, scheduling data are obtained. These include, but are not limited to, scheduled start date and time, duration, procedure, surgeon, and anesthesiologist. The schedulers at our institution reside both centrally in a surgical posting office and decentralized in physicians' offices (in the oral maxillofacial and organ transplant services). Because scheduling data do not directly enter the patient's medical record or roll directly into clinical documentation that must be reviewed and modified by a nurse, it is assumed that the risk for bias is minimal. Manipulation of case durations and scheduled start times is limited by system controls. The surgical posting office does have the ability to override system data (eg, for scheduled case duration, which is a by-product of historical averages), but this is not done without the approval from a supervisor.

Data from nursing documentation is under constant review by various clinicians to audit work. This process ensures data integrity and compliance and serves as a modest check and balance. Most of these data are objective, and although some bias may be present, this will most likely be minimal. The area of documentation that is most prone to bias is the "delay reason," because of its highly subjective nature and possible repercussions from management. Another factor for inaccurate delay reporting is the phenomenon of cascading delays (ie, when a delay early in the day causes delays in subsequent cases). By the time of the third or fourth delayed case, it is difficult to ascertain the cause other than to note that the previous case "ran over." During this series of delays, an entirely different cause of delay may happen in a specific case, but the reference point of a scheduled start time is lost, so that it is much more difficult to document a delay cause and duration. Of course, time stamps (in-room time minus scheduled start time) provide well documented and precise record of delay in minutes, but this does not qualify delay by reason type and provides no insights for root cause analysis.

The second threat to initiation of a system such as the one we developed lies in the general perceptions by staff and physicians. Many may feel that they are being spied upon or monitored, especially in areas in which no previous metrics existed. Others may find themselves out of their routine comfort zones. Underperforming staff who worry that they may be identified by the system may aggressively resist implementation of the new tools or work to undermine data integrity. Depending on the organization's structure, the open availability of data could result in punishment for individuals or a group rather than the intended promotion of positive departmental

and institutional change. Moreover, in environments in which competition for OR time is strong, surgical chiefs may be tempted to use data as a weapon to promote their own agendas.

The transparency of the data should alleviate some of these concerns. Team members should be able to see the data in which performance is being judged. In the past, data was obtained by someone walking around with a clipboard or in a back office recording data off charts, with limited or no ways to verify whether the data were true and accurate. With the drilldown features and different ways of organizing data for dashboard display, team members can easily view the raw data.

Conclusion

Strategic decisions made based on the management instinct have a lasting effect on the well-being of an organization. Management could benefit from the adoption of business intelligence tools that provide a quantifiable, validated alternative to instinct and ad hoc choices decision making.

Behavior and practice changes are central to achieving the objective of quality reports that drive efficiency. Too often, data are not integrated within the scope of daily practice. Acceptance of the importance of data must become a part of the culture of the organization. Graphical dashboards that present information down to the simplest, easiest-to-understand, and most accurate levels can compel this behavior change. Managers in the perioperative environment should seize the opportunity to integrate data into their organizations' cultures. Our research suggests that one promising approach is in Web-based tools that can be made for targeted audiences and adjusted by role, position, or location. The result can be total participation in quality improvement and constant feedback that provides long-term rewards in cost efficiencies, staff and physician satisfaction, and improved patient outcomes.

Acknowledgments

This project was supported by an OR of the Future grant from the Telemedicine and Advanced Technology Research Center. We would like to thank Dr Nancy Knight from the University of Maryland for her expert assistance in preparing this manuscript.

Reference

1. Weick KE, Sutcliffe KM. *Managing the Unexpected—Assuring High Performance in an Age of Complexity*. San Francisco, CA: Jossey-Bass; 2001.
2. Dexter F, Xiao Y, Dow AJ, Strader MM, Ho D, Wachtel RE. Coordination of appointments for anesthesia care outside of operating rooms using an enterprise-wide scheduling system. *Anesth Analg*. 2007;105:1701-1710.
3. Dexter F. Why calculating PACU staffing is so hard and why/how operations research specialists can help. *J Perianesth Nurs.* 2007;22:357-359.
4. Dexter F. Bed management displays to optimize patient flow from the OR to the PACU. *J Perianesth. Nurs*. 2007;22:218-219.
5. Dexter F. Operating room utilization: information management systems. *Curr Opin Anaesthesiol*. 2003;16:619-622.
6. Dexter F, Epstein RH, Traub RD, Xiao Y. Making management decisions on the day of surgery based on operating room efficiency and patient waiting times. *Anesthesiology*. 2004;101:1444-1453.
7. Macario A, Chow JL, Dexter F. A Markov computer simulation model of the economics of neuromuscular blockade in patients with acute respiratory distress syndrome. *BMC Med Inform Decis Mak*. 2006;6:15.
8. O'Sullivan CT, Dexter F, Lubarsky DA, Vigoda MM. Evidence-based management assessment of return on investment from anesthesia information management systems. *AANA J*. 2007;75:43-48.
9. O'Neill L, Dexter F. Tactical increases in operating room block time based on financial data and market growth estimates from data envelopment analysis. *Anesth Analg*. 2007;104:355-368.
10. Davenport TH. Competing on Analytics. *Harv Bus Rev.* January 2006;1-12.
11. Davenport TH, Harris JG. Competing On Analytics: The New Science of Winning. 1st ed. Boston, MA: Harvard Business School Press; 2007.
12. Chisholm M. The Twin Towers of BI Babel: Enterprise Architecture. BI Review, December 2007. http://www.bireview.com/issues/2007_42/10000440-1.html. Accessed January 10, 2008.
13. Beer S. Business Intelligence top priority of CIOs, February 2007. http://www.itwire.com.au/content/view/9906/53. Accessed January 10, 2008.
14. Pande PS, Neuman RP, Cavanagh RR. *The Six Sigma Way: Team Fieldbook*. New York, NY: McGrawHill; 2002.
15. Malik S. *Enterprise Dashboards: Design and Best Practices for IT*. Hoboken, NJ: John Wiley & Sons; 2005.
16. Treisman. Preattentive processing in vision. *Comput Vis Graph Image Process*. 1985;31:156-177.
17. Treisman A, Gormican S. Feature analysis in early vision: evidence from search asymmetries. *Psychol Rev.* 1988;95:15-48.

18. Washburne JN. An experimental study of various graphic, tabular, and textual methods of presenting quantitative material. *J Educ Psychol*. 1927;18:361-376.
19. Tufte ER. *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press; 1983.
20. Cleveland WS, McGill R. Graphical perception and graphical methods for analyzing scientific data. *Science*. 1985;229:828-833.
21. Montazemi AR, Wang S. The effects of modes in information presentation on decision-making: a review and meta-analysis. *J Manage Inf Syst*. 1988;5:101-27.
22. Feldman-Stewart D, Brundae MD, Zotov V. Further insight into the perception of quantitative information: judgments of gist in treatment decisions. *Med Decis Making*. 2007;27:34-43.
23. Meyer J, Shamo MK, Gopher D. Information structure and the relative efficacy of tables and graphs. *Hum Factors*. 1999;41:570-587.
24. Shneiderman B. Inventing discovery tools: combining information visualization with data mining. *Inf Vis*. 2002; 1:5-12.
25. Procedural Times Glossary of the AACD. http://www.aacdhq.org/Glossary.htm. Accessed January 10, 2008.
26. Charters KG. Nursing informatics, outcomes, and quality improvement. *AACN Clin Issues*. 2002;14: 282-294.
27. Dexter F, Ledolter J, Wachtel RE. Tactical decision making for selective expansion of operating room resources incorporating financial criteria and uncertainty in subspecialties' future workloads. *Anesth Analg*. 2005;100: 1425-1432.
28. Strum DP, Vargas LG, May J. Surgical subspecialty block utilization and capacity planning: a minimal cost analysis model. *Anesthesiology*. 1999;90:1176-1185.

A Research Portfolio for Innovation in the Surgical Environment

Gerald R. MOSES,[a, 1] PhD, Adrian E. PARK,[b] MD,
[a] *University of Maryland Medical Center, Baltimore MD*
[b] *Department of Surgery, University of Maryland Medical Center*

Abstract. The University of Maryland Medical Center and School of Medicine have sponsored a program of research targeted at the enabling of technologies for enhanced training, clinical effectiveness and patient safety. The pillars of this research included scientific approaches related to Informatics, Smart Image, Simulation and Ergonomics and Human Factors. The evolving research effort opened the door to a revised concept of basic surgical sciences that underpin training and performance in the operative environment.

Keywords. Surgery, training, innovation, surgical basic sciences

Background

The phrase, operating room of the future (ORF), has been used to describe the development of medical technology and the improvement of function and safety of the perioperative environment. The research program in the Department of Surgery at the University of Maryland has extended the meaning of the ORF to the study of functions and interactions of people, processes and technology producing a safe and efficient operating suite.

The Research Portfolio

For five years, the University of Maryland Medical Center and School of Medicine have sponsored a program of research targeted at the enabling of technologies for enhanced training, clinical effectiveness and patient safety. Initially, under the rubric of "The Operating Room of the Future" various pillars of research were established that proposed to advance the state of medicine, notably surgery. The pillars included scientific approaches related to Informatics, Smart Image, and Simulation. The evolving research effort opened the door to a revised concept of basic surgical sciences that underpin training and performance in the operative environment.

Developments led to two important changes; the adoption of a new mantra, Innovation in the Surgical Environment, to replace the Operating Room of the Future; and the addition of another research pillar, that of Ergonomics and Human Factors.

[1] Corresponding Author: Gerald Moses, MASTRI Center, Division of General Surgery, University of Maryland Medical Center, 22 S. Greene Street, Baltimore, MD 21201; gmoses@smail.umaryland.edu

Progress has been achieved in each of the pillars of research, as reported at a recent annual conference that sought to apply lessons learned form the high-stakes environments of aviation and astronautics to the practice of surgery.

Research Pillars

The medical informatics pillar includes a Perioperative Scheduling Study, a study of workflow around performance indicators in the peri-operative environment and building a graphical dashboard to allow data mining and trend analysis of operating indicators. The surgical simulation pillar entails both physical and cognitive simulation for training with emphasis upon laparoscopic surgery. A third pillar is entitled “Smart Image” in which we are seeking to push the boundaries of real time deformable image registration with a goal of performing the 1st fully smart image guided laparoscopy. A recently added pillar of ergonomics and human factors addresses the impact of stress movements and position upon the surgeon performing minimally invasive or “open” procedures.

Informatics: Workflow and Operations Research for Quality (WORQ)

The Perioperative Scheduling Study is looking at how using post-operative destination information during the process of surgery scheduling can influence congestion in post-operative units such as intensive care units (ICUs) and intermediate care units (IMCs), which lead to overnight boarders in the post-anesthesia care unit (PACU). We have developed a mathematical congestion evaluation model for evaluating congestion in post-operative units, including ICUs, IMCs, and floor units. This model requires data about post-operative destinations and length-of-stay distributions for different types of surgeries. We have analyzed data about cardiac surgeries from two years and have analyzed UMMC financial records for all of the surgical cases for fiscal year 2007. We have developed an algorithm for predicting bed requirements based on the surgical schedule and have conducted a preliminary study comparing these predictions to other prediction methods for two units. The preliminary results show that the new bed requirements prediction method is more accurate.

Informatics: Operating Room Glitch Analysis (OGA)

The OGA project, focusing on institutional learning, is looking at the workflow around performance indicators in the peri-operative environment and building a graphical dashboard to allow data mining and trend analysis of operating indicators.

We have integrated into the data architecture a javascript based bubble chart that provides several interactive features to allow thorough data discovery. The bubble chart can play over time to see how the size of the bubbles change, which relates to the number of cases performed, as well as their x and y axis location. The x and y axis can represent delay duration, actual procedure time, scheduled procedure time, or turnover time. The bubble can also be tagged to provide a contrail to show performance over time. Figure 1 indicates an analysis of service delay as related to average length of surgical procedure.

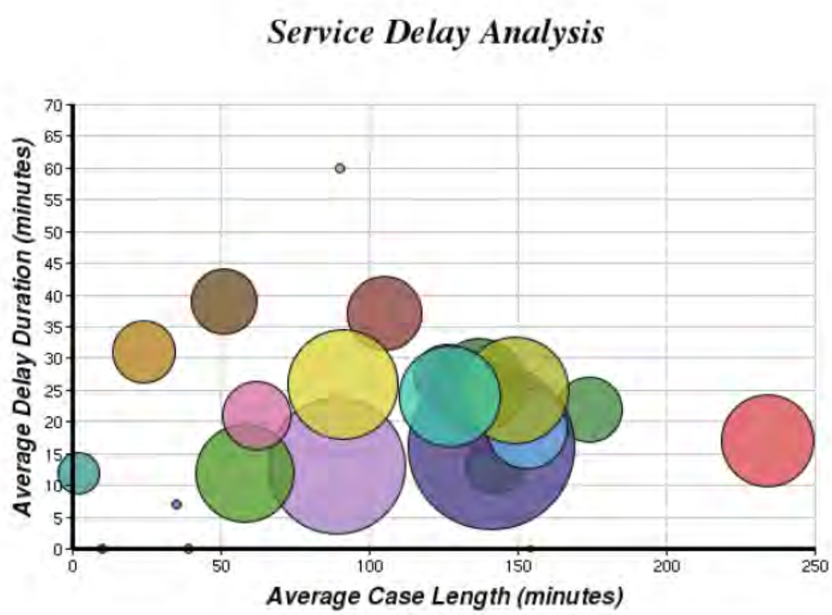


Figure 1. Service delay as related to average length of surgical procedure.

Informatics: Video Summarization of Key Events in Surgery

The technique of summarization is used when confronted with the task of gleaning succinct information from large amounts of data. For example, our national intelligence services use both machine and human analysis to prepare the daily Intelligence Summary for the President. A similar challenge is presented to those who train surgeons using a vast archive of surgical video. A key element in teaching is the extraction of the right video event to make the critical point to surgical trainees.

Recent decades have seen an increasing use of VR and simulation aids in surgical training. The typical approach is to use sensors to capture the kinematics of the tools, as well as force/torque measures. One thread of work directly analyzes these measurements to construct Markov Models that describe the state and transitions for a surgical procedure, and it is then shown that the transition probabilities between states are different at different levels of expertise.

An alternative approach is for an expert to look at the video of the surgical procedure (or training), identify key steps/events (either done well or incorrectly), and then judge the skill level of the performer. This approach can bring to bear the expert's knowledge and intuition of the complex interaction between tools, movements, organs, cutting planes etc. The drawback however is that it requires the review of a video that can be very time-consuming. We propose to address this problem by developing techniques to automatically identify key scenes/events in a video of laparoscopic surgeries.

Simulation

We are conducting multiple studies of the effects of physical box trainers, virtual reality (VR) trainers, and mixed modality training for acquiring laparoscopic surgery skills. These studies support the actions and operations of the Maryland Advanced Simulation Training, Research and Innovation (MASTRI) center. Additionally, we are developing a cognitive simulator and building knowledge representations based on ontology of focused human anatomy/physiology to emulate the surgical/clinical experience. The cognitive simulator, the Maryland Virtual Patient, has been developed by construction of a computational model of the cognitive agent, and by testing the goal- and plan-based reasoning component and its interaction with the interoceptive and language perception modules and verbal, mental and physical action simulation modules.

We have continued to work on the natural language substrate of the system, concentrating on enhancements required for processing dialog (not expository text). Further, we have implemented an enhanced microtheory of indirect speech acts, and continued working reference resolution algorithms.

The research work encompasses work targeted upon the acquisition of ontology and lexicon knowledge., and improvement of the DEKADE user interface. The current version of the cognitive simulation system includes multiple scenarios of physician-patient interface related to LERD/GERD patient conditions.

Smart Imaging

Surgical practice is considered among the most complex and difficult fields. That no two patients are exactly alike is one of the challenges that make it so. Anatomic and physiologic differences make each case unique. In surgery, these variations can complicate an operation; the discovery of unexpected anatomical variations often requires a surgeon to stray from standard, well-practiced techniques to attempt a novel approach to the procedure. With novelty comes a reduced margin of safety. This situation is exacerbated by a trend toward further physical separation between the patient and interventionalists (e.g., surgeons, endoscopists, radiologists) and a greater dependence on an image of the patient's (target) anatomy to effect therapy or establish a diagnosis.

"Smart image," as we have defined it, refers either to the process of extracting elements from an environment and imparting them to an image or to acquiring elements from within a scene and enhancing them. The result in either case is a more meaningful visualization of the operative field. Although many applications exist within this definition, Maryland's smart image team is working toward performing the first laparoscopic surgery guided completely by smart image.

Typically in laparoscopic procedures, diagnostic imaging—including x-rays, computerized tomography (CT), and magnetic resonance imaging (MRI) scans—can provide a preview of patient physiology. Often, however, these diagnostic images are in a static format that does not allow the care provider to interact meaningfully with the information the images contain. Current advances in smart imaging can be used to improve patient safety by providing the caregiver with a more interactive experience. A set of two-dimensional (2D) slices of a CT scan can be transformed into a three-

dimensional (3D) computer model so that surgeons can preview a realistic view of the patient's anatomy before an operation. This type of smart imaging provides an interactive "fly-through" view that allows the surgeon to explore the anatomy in detail.

With advances in computing power, these previews could be mapped more realistically to interactive simulators that would permit rehearsal of a surgical procedure that might include attempts at novel approaches before surgery begins. During real surgery, these smart diagnostic images could be integrated into the surgeon's actual view of the patient.

We are working toward matching the minimally invasive surgeon's video view of the surface anatomy with computer-generated models from Digital Imaging and Communications in Medicine (DICOM) data sets. Such imaging could provide the surgeon with real-time "x-ray vision" during the operation. Thus, the underlying structure, such as the position of a tumor beneath the surface of a larger anatomic structure or blood vessels within the liver, could be seen. Vessels could be contrast-enhanced in a single, high-resolution CT scan before the surgery. Then, during surgery, low-dose/low-resolution CT scans could be used to transform the high-resolution CT image to match the movement of the patient's anatomy during surgery. This would allow intraoperative visualization of anatomy that retains the enhanced contrast vessels, a unique ability that is not possible at present.

CT scans can provide enhanced intraoperative visualization of deep structures far superior to that of laparoscopes. However, the use of continuous CT exposes the patient and surgeon to a radiation level that remains a concern. Therefore, a major thrust of our work is to design, develop, and test several dose-reduction strategies and to incorporate these into our proposed continuous CT-guided surgical navigation system. Our preliminary work suggests that our strategies would allow us to lower the net radiation exposure to the patient to levels commonly viewed as safer in cardiac catheterization and interventional radiology procedures. In the long term, we also propose using telemanipulators to remove surgeons from the CT room and thereby shield them entirely from radiation exposure while they are performing the procedure.

Ergonomics and Human Factors

Recently, a fourth pillar was added to our research portfolio, that of Ergonomics and Human Factors. These are two related branches of study that examine the relationship between people and their work environment. Ergonomics often focuses on the physical environment and the human body, while human factors center more on the cognitive aspects of performance. The same ergonomics and human factors techniques credited with making industrial processes safer and more efficient can be applied to the analysis and improvement of OR operations. Tools, such as video analysis and motion tracking, can be used to analyze current practices, identify inefficiencies and dangers, develop solutions, and measure improvement. "Best practices" to maximize safety and efficiency can be developed based on empirical data.

Our discussion of workflow to this point has taken a macro or panoramic view; for example, how might we most effectively track and bring together the people and assets necessary to ensure that a patient's surgical experience is safe and efficient. Through human factors and ergonomics, we have the ability to focus on a more micro-level

analysis, such as measurements of surgeon/instruments interface and how the physical interface between the surgeon and the patient could be improved.

In the future, OR workspace layout would be optimized through ergonomic data and human factors analysis, and this optimization would lead to the establishment of “best practices” for an array of surgical operations. Proper layout would reduce risks of infection, speed operations, and reduce fatigue of surgeons and staff, all elements that could contribute to a reduction in adverse events and improved patient safety.

Future Vision of the Operating Room Environment

Well-trained care providers, who have reached a level of proficiency on realistically simulated patients, are supported by an array of smart technology enabling surgical procedures to be performed in an ever safer environment. Cases start on time with all team members informed of the goals and possible trouble spots of each operation. Contingency plans are in place for dealing with anticipated complications. The smart environment checks that all required equipment and people are present and cross-checks drugs and blood products brought into the room, ensuring patient compatibility in terms of allergies and blood type. Surgeons do not have to fight fatigue and discomfort during surgery, as the layout of the surgical workspace is ergonomically correct. Thus, the time and effort needed to perform surgery is minimized and improvement of both technique and outcomes is realized.

A New Set of Basic Surgical Sciences

The potential of surgical care in the future can be realized by incorporating into the training of surgeons a new set of basic surgical sciences, those of advanced imaging, informatics systems, simulation and ergonomics and human factors. These do not replace the well established scientific bases of anatomy, physiology, pathology and related areas of study. Rather they add a vital underpinning to the knowledge and expertise required of future practitioners.

References

[1] Shekhar, R., Dandekar, O., Kavic, S.,George, I., Mezrich, R., and Park, A. "Development of continuous CT-guided minimally invasive surgery," in Medical Imaging 2007 Visualization and Image-Guided Procedures, San Diego, CA, USA, 2007, pp65090D-8.

[2] Shetye, A. and R. Shekhar, "A statistical approach to high-quality CT reconstruction at low radiation doses for real-time guidance and navigation," in Medical Imaging 2007: Physics of Medical Imaging, San Diego, CA, USA, 2007, pp. 65105U-11.

[3] Lee G, Kavic SM, George IM, Park AE (2007) MIS surgical ergonomics: Future Trends, Annual conference of Medicine Meets Virtual Reality (MMVR), Long Beach, CA.

[4] Moses G.R., Seagull FJ, George I.M.and Park A.E. The MASTRI Center – Medical Simulation for Skill Acquisition. Proceedings of MODSIM World Conference 2007

Surg Endosc (2008) 22:1087–1092
DOI 10.1007/s00464-007-9664-3

Development of a more robust tool for postural stability analysis of laparoscopic surgeons

Gyusung Lee · Adrian E. Park

Received: 30 May 2007 / Accepted: 9 October 2007 / Published online: 20 November 2007

Abstract
Background Physical difficulties experienced by surgeons performing minimally invasive surgery (MIS) are being given extensive attention by ergonomic researchers. Postural stability, not commonly addressed, is our prime focus. Center of pressure (COP) alone is used in the few existing postural stability studies. Using COP, we previously correlated postural stability to instrument type, task difficulty, and skill level. This study, including center of mass (COM), sway area analysis, and what we uniquely term postural stability demand (PSD), extends our investigation.
Methods Six surgeons from different experience levels were recruited to complete three fundamentals of laparoscopy (FLS^{TM}) tasks. Standing on two force plates, participants performed each task as a motion capture system recorded body movements. An ellipse was created for sway area analysis of COP, the point where the ground reaction force was located, and COM, the point at which body mass was concentrated. PSD was defined as the mean distance between the COP and COM locations in the anterior–posterior (A–P) or medial–lateral (M–L) directions. Postural parameters and performance time were correlated.
Results COM and COP sway areas positively correlated with pegboard transfer performance time ($r = 0.928$, $p < 0.05$; $r = 0.864$, $p < 0.05$) and also with circle-cutting performance time ($r = 0.858$, $p < 0.05$; $r = 0.779$, $p = 0.06$). However, COM and COP sway areas negatively correlated with endo-loop placement performance time ($r = -0.925$, $p < 0.05$; $r = -0.935$, $p < 0.05$). These results indicate unique postural controls based on skill level. During all tasks, PSD in the A–P direction strongly correlated with performance time ($r = 0.829$, $p < 0.05$; $r = 0.913$, $p < 0.05$; $r = 0.880$, $p < 0.05$), indicating that less-skilled participants experienced increased postural demands.
Conclusions This study demonstrated that variance in postural adjustments, as evidenced by sway area analysis, correlate to skill level and individual task. Strong correlation between PSD and performance time shows potential as a predictor of skill levels. Combining COM, COP, and PSD data produces a more robust analytic tool for identifying postural adjustments that can be correlated with skill level.

Keywords Laparoscopy · Ergonomics · Postural analysis · Center of pressure (COP) · Center of mass (COM) · Force plate

G. Lee · A. E. Park (✉)
Division of General Surgery, Department of Surgery, School of Medicine, University of Maryland, 22 South Greene Street Rm. S4B14, Baltimore, Maryland 21201
e-mail: apark@smail.umaryland.edu

G. Lee
e-mail: glee@smail.umaryland.edu

The benefits of minimally invasive surgery (MIS) to patients, including shortened hospital stays, shortened recovery times, and less scarring, are often cited. Performing a laparoscopic surgical procedure, however, often places physical demands on MIS surgeons that differ substantively and dramatically from those during the performance of open surgery. Several MIS components, including long shaft instruments, access ports, and the endoscope image display system, cause numerous challenges for surgeons learning and practising laparoscopy. Unfavorable physical ergonomics, high cognitive demand, and limited haptic and visual information are among these challenges [1, 2].

Previous studies in surgical ergonomics have investigated how these difficult challenges might be minimized. New instrument designs [3, 4] as well as operating room (OR) environment factors, including display location and height [5, 6], instrument placement and task alignment [7–9], and operating table height [10, 11], have been evaluated for their effect on task performance and on surgical ergonomic variables such as joint angles and muscle activation. In many of these studies, the upper-body joint rotations captured by motion analysis systems in combination with muscle activation data recorded by electromyography (EMG) systems have been commonly used as assessment tools to quantify physical workload and muscular fatigue.

The human body always requires continuous active control to maintain proper balance. The body, which has been characterized as an unstable balance system, has two-thirds of its mass located at two-thirds of its height above ground level [12]. For this reason, posture during quiet standing, often described as an inverted pendulum, appears to be static but actually relies on dynamic control [13]. It is well known that maintaining correct posture is a very important ergonomic factor credited not only with minimizing physical discomfort but also with improving task performance [14–16].

In various industrial workplaces, ergonomic assessments have been undertaken to quantify postural stress/discomfort levels in attempts to describe optimal work postures [17, 18]. With healthcare workers as subjects, posture studies have been carried out in increased numbers, though it has been reported that their focus has primarily been low-back pain problems experienced by hospital nurses [19].

Maintaining good postural stability has been considered very important for surgeons performing laparoscopic tasks. Postural balance control can only be assessed using data acquired through force plates, which record the amplitude and direction of the ground reaction force and the location of the center of pressure (COP). Force plates have commonly been used in biomechanical and neuroscience studies to assess postural data during quiet standing [20], perturbed standing [21], and functional standing or walking [22, 23]. However, only a few studies have applied this technology in surgical ergonomic investigations. Berguer et al. used a single force plate to compare surgeons' postures during laparoscopic and open surgical procedures [24]. This study showed that surgical posture during laparoscopic procedures was less dynamic, signified by significantly reduced range of motion (ROM) of the COP. Gillette et al. have found that COP excursions, which were defined by outer boundaries, increased with more difficult training tasks [25].

In this study, we introduce, as an extension of our previous finding, a more robust way to analyze force plate data. To calculate sway areas, we constructed an ellipse that covered 95% of COP excursions. We postulated that this assessment tool would be less sensitive to erroneous data outliers and would provide more details of postural control. We also used a new variable that we term postural stability demand (PSD) to explain quantitatively how the center of mass (COM) is related to COP excursions.

Materials and methods

This institutional review board (IRB)-approved study was performed in the Surgical Ergonomics Laboratory at the Maryland Advanced Simulation, Training, Research, and Innovation (MASTRI) Center of the University of Maryland, School of Medicine. Six subjects possessing different levels of MIS experience were recruited from the Department of Surgery at the University of Maryland, School of Medicine to complete pegboard transfer, endo-loop placement, and circle-cutting tasks from the fundamentals of laparoscopy (FLS™) skill set. The FLS™ modules were developed by the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) to assess technical surgical skills. All subjects volunteered for this study and signed an informed consent. Endoscope images from a 0° scope were displayed on a standard cathode-ray tube (CRT) monitor positioned at eye level in front of the participants. Directly before each trial, participating surgeons watched an instructional FLS™ video clip that detailed the requirements of each task. Thirty-nine 9.5 mm retroreflective sphere markers were attached to each participant according to marker placement designated by Plug-in-Gait™ (Lake Forest, CA). The markers were placed on (1) bands around the head, wrists, and knees, (2) a custom-made vest and waist belt, (3) the skin of the arms, and (4) the shoulder, thighs, and shanks of the participants' medical scrubs. The elastic bands and vest over the scrub suit of participating subjects allowed the markers to be securely attached to anatomical landmarks. A ViconPeak™ (Lake Forest, CA) motion capture system consisting of 12 high-speed, high-resolution, infrared, digital cameras tracked the markers and reconstructed body segment movement in three-dimensional (3-D) space. Location data of each marker were sampled at 100 Hz. All tasks were performed at a trainer box that rested on a height-adjustable table. In calculating the optimal table height, we relied on the findings of Berquer et al., which define the optimal height as that at which the laparoscopic instrument handles are at elbow height or 10 cm below elbow height [26]. In our study the supervising ergonomist adjusted the table height accordingly. The trainer station was located in front of two force plates (Advanced Mechanical Technology Inc., Watertown, MA). Each task began when a verbal signal was given. Performance time was measured during each task and used as an indirect measure of surgical skill level.

During each task, participants stood with one foot on each force plate, which sampled ground reaction forces and moment data at 100 Hz. A fourth-order Butterworth low-pass filter with a cut-off frequency of 6 Hz was applied to the raw force plate data to remove high-frequency noise.

We used two parameters—COM and COP—as the determinants for postural stability analysis. The entire body COM for each subject was calculated by integrating anthropometric data and body kinematic data, the latter being obtained from motion analysis. COP data was measured by two force plate systems. Calculation of the location of the COP for the left and the right leg was acquired by having each subject perform each FLS™ task while standing on two separate force plates with one foot on each. Overall COP ($COP_{overall}$) was defined as the weighted average of each force plate COP based on vertical ground reaction forces (Fig. 1).

$$COP_{overall} = COP_{Left}\frac{R_{VL}}{R_{VL} + R_{VR}} + COP_{Right}\frac{R_{VR}}{R_{VL} + R_{VR}},$$

where COP_{Left} and COP_{Right} are calculated under the left and right foot, respectively, and R_{VL} and R_{VR} are the magnitude of the vertical ground reaction forces from each foot. From here on, when COP is mentioned, it is overall COP that is being referred to.

Traditionally COM and COP excursions are defined by sway amplitude in both the medial–lateral (M–L) and anterior–posterior (A–P) directions. However, we characterized these excursions by sway area. To do this we used the principal component analysis (PCA) technique, also known as singular value decomposition (SVD), to calculate an ellipse whose area covered 95% of the data points of the COM and COP excursions. Figure 2 shows how such an ellipse is fitted to force plate data. This technique provides the lengths of the major axis (L_1), the minor axis (L_2), and the angle of major axis (θ). We estimated the sway area as the following:

$$Sway\,area = L_1 \times L_2 \times \pi$$

To define PSD we used the two-dimensional (2-D) coordinates of the COP and COM in the A–P and M–L directions and excluded the height of the COM coordinate

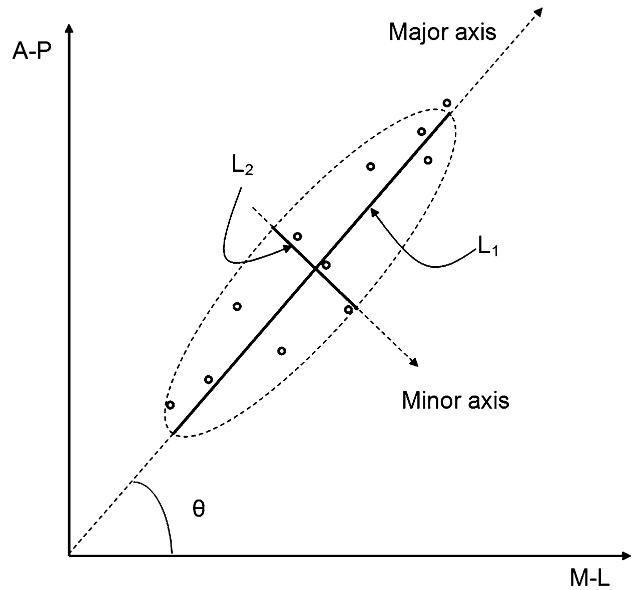


Fig. 2 Ellipse created by using principal component analysis (PCA) was fitted to simplified data points of COP excursions. The lengths of the major and minor axes are L_1 and L_2, respectively. The angle of the major axis, θ, is the sway angle

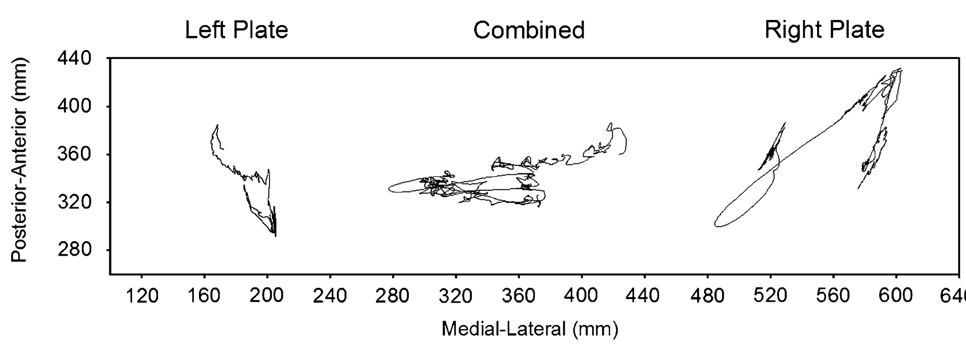
(Fig. 3). The PSD variable indirectly represents postural instability, which should be controlled to maintain good balance. Postural stability demand was calculated as the mean distance between the COM and COP locations in the A–P or M–L directions. During perfect, quiet, upright standing, PSD is close to zero.

$$PSD_{A-P} = COM_{A-P} - COP_{A-p}$$

$$PSD_{M-L} = COM_{M-L} - COP_{M-L}$$

Three postural parameters—COM, COP, and PSD—were correlated with task performance time, an indirect measure of surgical skill, to investigate whether unique balance controls as evidenced by a variety of COM and COP sway areas and PSD distances were required for performance of different FLS™ tasks or by surgeons with differing surgical skill levels. We performed all data analysis through custom-developed programs written in the Matlab™ (Mathwork, Natick, MA) language.

Fig. 1 Example of COP excursion data recorded from force plates: left plate COP trajectories; combined overall COP using weighted average based on vertical ground reaction forces recorded by each force plate; right plate COP trajectories



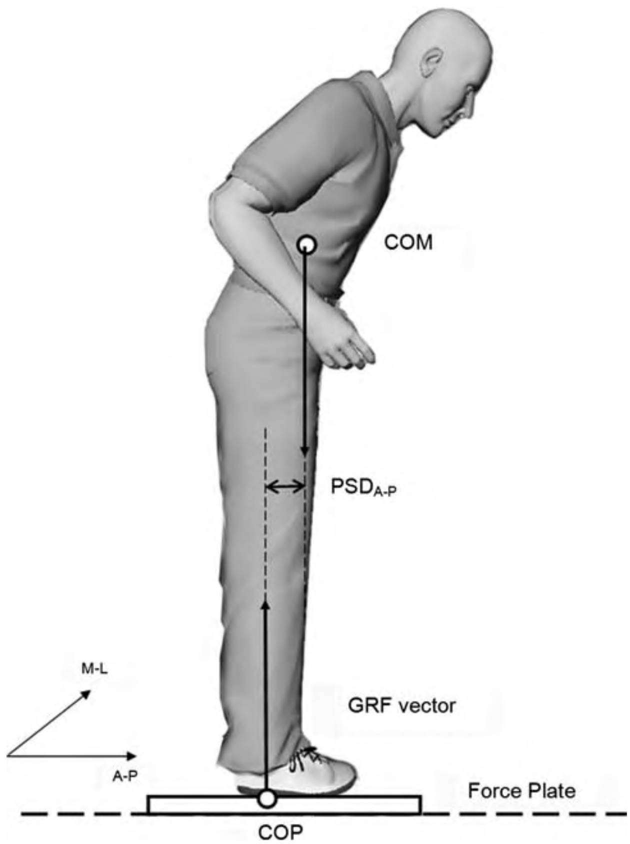


Fig. 3 Illustrations of the postural parameters COM, COP, and PSD. COM is the point where all the mass of a body is concentrated, and COP locates the acting point of the resultant ground reaction force (GRF). PSD is defined here as the mean distance between the COM and COP locations in the A–P and M–L directions

Results

Figure 4 shows the correlation between the sway area of COM excursion and the performance time for each FLS^{TM} task. During the pegboard transfer and circle-cutting tasks, sway areas were positively correlated with performance time ($r = 0.928$, $p < 0.05$; $r = 0.858$, $p < 0.05$), indicating that those participants whose performance times were shorter exhibited more static posture. During the endo-loop placement task, however, sway areas negatively correlated with performance time ($r = -0.925$, $p < 0.05$), indicating that those with shorter performance time showed more dynamic body sway. From these results we postulate that the surgeons who performed faster were using task-specific and skill-related postural sway control strategies.

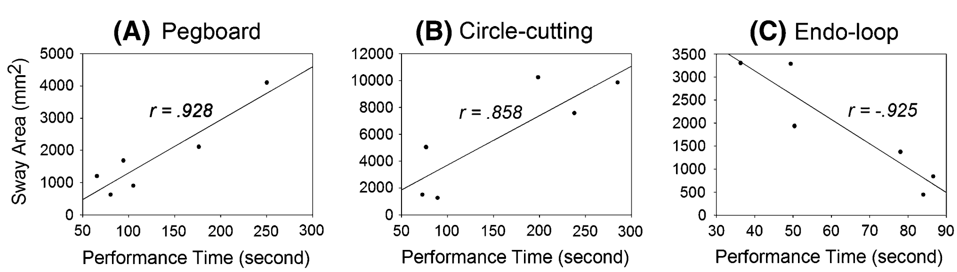
Similar results obtained from COP sway area analysis are shown in Fig. 5. Positive correlation between sway area and performance time during pegboard transfer and circle-cutting tasks ($r = 0.864$, $p < 0.05$; $r = 0.779$, $p = 0.06$) and negative correlation during endo-loop placement task ($r = -0.935$, $p < 0.05$) were found for surgeons whose performance times were shorter. From the fact that our sway area analysis of COM and COP showed very similar patterns, we infer that the upper-body movements of the surgeon are correlated with balance control provided by the ankle and that the upper body and ankles interact with each other so that proper postural control can be achieved during each task.

Figure 6 shows the results of our correlation analysis between PSD and performance time. PSD in the A–P direction (PSD_{A-P}) was strongly correlated with performance time during all three FLS^{TM} tasks ($r = 0.829$, $p < 0.05$; $r = 0.913$, $p < 0.05$; $r = 0.880$, $p < 0.05$). PSD in the M–L direction, however, did not show any correlation with performance time. Even though postural balance control varied with each task, PSD was consistently higher for participants whose task performance time was longer, suggesting that this group displayed a high degree of postural instability.

Discussion

Laparoscopic surgeons accommodate their motions to facilitate their movements with a variety of instruments based on the static locations of trocars. The performance of MIS, therefore, consists often of awkward movements, including uncomfortable arm extensions and leaning postures. Fatigue and other short-term effects and muscular injury in addition to other long-term effects result from this type of continuous positioning and repositioning. To account for the dynamic upper-body movements that surgeons performing laparoscopy are required to make while standing static for a considerable time span, we theorized a difference in posture among novice and expert surgeons. That skill level variance could accurately be described we

Fig. 4 Correlations between the COM sway area and performance time during: **(A)** pegboard transfer, **(B)** circle-cutting, and **(C)** endo-loop placement tasks. The correlation coefficient (r) is indicated in each plot



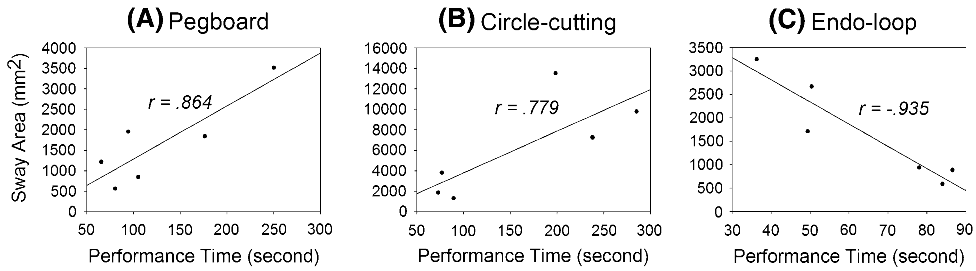


Fig. 5 Correlations between COP sway area and performance time during: **(A)** pegboard transfer, **(B)** circle-cutting, and **(C)** endo-loop placement tasks. The correlation coefficient (r) is indicated in each plot

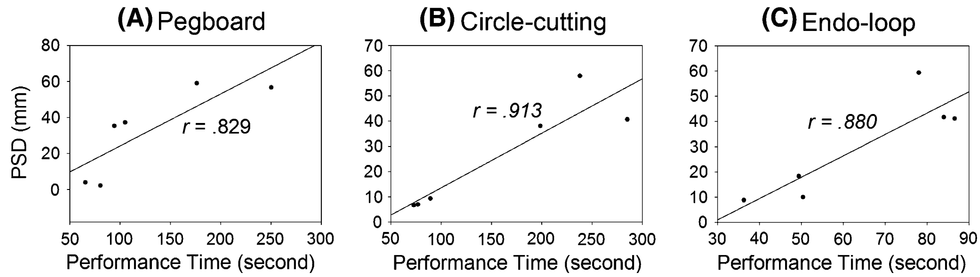


Fig. 6 Correlations between PSD and performance time during: **(A)** pegboard transfer, **(B)** circle-cutting, and **(C)** endo-loop placement tasks. The correlation coefficient (r) is indicated in each plot

proposed by using sway area analysis and PSD as determinants of the relationship between upper-body movement and balance and control as exerted at the foot level.

Our proposal as investigated in this study is a continuation of our previous reports on surgical postural analysis in which we detailed changes in surgeons' postures during the performance of three FLS^{TM} tasks: pegboard transfer, circle-cutting, and endo-loop placement. We first investigated COM excursions as exhibited by an experienced MIS surgeon with a wrist complication. This case study showed that postural instability as indicated by higher COM excursions does not necessarily correlate to poor performance since the surgeon being studied made compensatory arm movements to minimize wrist flexion and also exhibited strategic movements that while seeming to signal postural instability actually proved necessary to achieve successful task performance [27].

In another study, we analyzed the postural control strategies used by surgeons with different experience levels as they performed each of three FLS^{TM} tasks [28]. We correlated COP excursions, defined as ROM in the A–P and M–L directions to performance time. Our results showed that surgeons with shorter performance time had smaller sway in the A–P direction during pegboard transfer and in the M–L direction during circle-cutting task. Additionally, during endo-loop placement task, those who performed faster showed higher COP excursions in both the A–P and M–L directions. We concluded that during each task surgeons whose performance time was shorter demonstrated task-specific, skill-related postural controls.

Through our use of sway area analysis in our current study we have corroborated the unique balance control mechanisms identified earlier. In our current study, variance in postural balance adjustments as shown through sway area analysis could be correlated not only to individual task but also to skill level.

Use of the ellipse in sway area analysis also confers other advantages. We find that sway area analysis is preferable to overcome the limit of ROM analysis. ROM—calculated as the difference between maximum and minimum values—can be incorrectly biased by the existence of extreme outliers, which may be erroneous. The ellipse we created using 95% confident intervals made our sway area analysis less sensitive to outliers. Additionally, by using an ellipse in our sway area analysis we were able to determine the existence of directional sway patterns using both the sway angle, which is the angle between the major axis and the x-axis, and the ratio between the major axis and the minor axis. We know that the sway is directional if the ratio is larger than one (i.e., the major axis is longer than the minor axis). Directional data acquired through sway area analysis, which is less sensitive to outliers, permits the nature of postural control to be explained more precisely.

We added PSD in addition to sway area analysis for the express purpose of establishing and investigating the effect exerted by the relationship between upper-body joint movement (COM) and ground reaction force (COP) on postural control. According to previous biomechanics research, COP excursion is believed to specify how the ankle joint is controlled to maintain stability while COM movement is believed to show how effective this ankle control is [29, 30].

COP measurement has been used widely in biomechanics research because it is relatively simple to obtain data by using force plates. Useful posture information is provided by COP analysis, though it alone provides no data about the upper body movement that may be the cause of postural sway observed in COP excursions.

We propose the use of PSD, calculated from both COM and COP data, as a more sound and accurate description of the relationship between these two points. Our results, for instance, showed that surgeons whose performances were slower had higher PSD in the A–P direction as they performed FLS™ tasks that required variations of postural balancing. From the results of this study, we also propose the potential of PSD as an indirect predictive measure of surgical skill levels.

In conclusion, we demonstrated in this study that multiple postural assessments, in this case COM and COP using sway area analysis and PSD, serve as a more robust analytic tool to identify patterns of postural balancing control unique to specific laparoscopic tasks and to determine how this control varies with different surgical levels as indirectly represented by performance time. The knowledge that results from multiple measurements provides more detailed, sophisticated descriptions of postural controls used by individual surgeons. The resultant comprehensive data also provides strategic insight into how postural control is influenced by level of experience and contributes useful information that may assist in developing universal surgical ergonomic guidelines.

Acknowledgements This study was supported by a grant from the US Army Medical Research and Materiel Command (USAMRMC) and equipment was provided in part by US Surgical. The authors acknowledge the thoughtful assistance of Rosemary Klein in the editing of this manuscript.

References

1. Berguer R, Forkey DL, Smith WD (1999) Ergonomic problems associated with laparoscopic surgery. Surg Endosc 13:466–468
2. Carswell CM, Duncan C, Seales WB (2005) Assessing mental workload during laparoscopic surgery. Surg Innov 12:80–90
3. Matern U, Kuttler G, Giebmeyer C, Waller P, Faist M (2004) Ergonomic aspects of five different types of laparoscopic instruments handles under dynamic conditions with respect to specific laparoscopic tasks: An electromyographic-based study. Surg Endosc 18:1231–1241
4. Van Veelen MA, Meijer DW, Uijttewaal I, Goossens RHM, Snijder CJ, Kazemier G (2003) Improvement of the laparoscopic needle holder based on new ergonomic guidelines. Surg Endosc 17:699–703
5. Huber JW, Taffinder N, Russell RCG, Darzi A (2003) The effects of different viewing conditions on performance in simulated minimal access surgery. Ergonomics 46:999–1016
6. Smith WD, Berguer R, Nguyen NT (2005) Monitor height affects surgeons' stress level and performance on minimally invasive surgery tasks. Stud Health Technol Inform 111:498–501
7. Emam TA, Hanna GB, Kimber C, Dunkley P, Cuschieri A (2000) Effect of intracorporeal-extracorporeal instrument length ratio on endoscopic task performance and surgeon movements. Arch Surg 135:62–65
8. Emam TA, Hanna G, Cuschieri A (2002) Ergonomic principles of task alignment, visual display, and direction of execution of laparoscopic bowel suturing. Surg Endosc 16:267–271
9. Emam TA, Hanna G, Cuschieri A (2002) Comparison of orthodox vs. off-optical axis endoscopic manipulations. Surg Endosc 16:401–405
10. Berguer R, Smith WD, Davis S (2002) An ergonomic study of the optimum operating table height for laparoscopic surgery. Surg Endosc 16:416–421
11. Van Veelen MA, Kazemier G, Koopman J, Goossens RHM, Meijer DW (2002) Assessment of the ergonomically optimal operating surface height for laparoscopic surgery. J Laparoendosc Adv Surg Tech 12:47–52
12. Winter DA (1995) Human balance and posture control during standing and walking. J Biomech 3:193–214
13. Gage WH, Winter DA, Frank JS, Adkin AL (2004) Kinematic and kinematic validity of the inverted pendulum model in quiet standing. Gait Posture 19:124–132
14. Bhatnager V, Drury CG, Schiro SG (1985) Posture, postural discomfort, and performance. Hum Factors 27:189–199
15. Liao MH, Drury CG (2000) Posture, discomfort and performance in a VDT task. Ergonomics 43:345–359
16. Van Wely P (1970) Design and Disease. Appl Ergon 1:262–269
17. Corlett EN, Bishop RP (1976) A technique for assessing postural discomfort. Ergonomics 19:175–182
18. Keyserling WM (1986) A computer-aided system to evaluate postural stress in the workplace. Am Ind Hyg Assoc J 47:641–649
19. Kant IJ, de Jong LC, van Rijssen-Moll M, Borm PJ (1992) A survey of static and dynamic work postures of operating room staff. Int Arch Occup Environ Health 63:423–428
20. Masani K, Popovic MR, Nakazawa K, Kouzaki M, Nozaki D (2003) Importance of body sway velocity information in controlling ankle extensor activities during quiet stance. J Neurophysiol 90:3774–3782
21. Rietdyk S, Patla AE, Winter DA, Ishac MG, Little CE (1999) Balance recovery from medio-lateral perturbations of the upper body during standing. J Biomech 32:1149–1158
22. Halliday SE, Winter DA, Frank JS, Patla AE, Prince F (1998) The initiation of gait in young, elderly, and Parkinson's disease subjects. Gait Posture 8:8–14
23. Karst GM, Venema DM, Roehr TG, Tyler AE (2005) Center of pressure measures during standing tasks in minimally impaired persons with multiple sclerosis. J Neurol Phys Ther 29:170–180
24. Berguer R, Rab GT, Abu-Ghaida H, Alarcon A, Chung J (1997) A comparison of surgeons' posture during laparoscopic and open surgical procedures. Surg Endosc 11:139–142
25. Gillette JC, Quick NE, Adrales GL, Shapiro R, Park AE (2003) Changes in posture mechanics associated with different types of minimally invasive surgical training exercises. Surg Endosc 17:259–263
26. Berquer R, Smith WD, Davis S (2002) An ergonomic study of the optimum operating table height for laparoscopic surgery. Surg Endosc 16:416–421
27. Lee G, Kavic SM, George IM, Park AE (2007) Postural instability does not necessarily correlate to poor performance: Case in point. Surg Endosc 21:471–474
28. Lee G, Kavic SM, George IM, Park AE (2006) Correlation between postural stability and performance time during fundamentals of laparoscopic surgery (FLS) tasks. Br J Surg 93(Suppl):S206
29. Benvenuti F, Mecacci R, Gineprari I, Bandinelli S, Benvenuti E, Ferrucci L, Baroni A, Rabuffetti M, Hallett M, Dambrosia JM, Stanhope SJ (1999) Kinematic characteristics of standing disequilibrium: Reliability and validity of a posturographic protocol. Arch Phys Med Rehabil 80:278–287
30. Nault ML, Allard P, Hinse S, Le Blanc R, Caron O, Labelle H, Sadeghi H (2002) Relations between standing stability and body posture parameters in adolescent idiopathic scoliosis. Spine 27:1911–1917

Review Article

Surgical Innovation
Volume 14 Number 3
September 2007 153-167

10.1177/1553350607307956
http://sri.sagepub.com
hosted at
http://online.sagepub.com

Methodological Infrastructure in Surgical Ergonomics: A Review of Tasks, Models, and Measurement Systems

Gyusung Lee, PhD, Tommy Lee, MD, David Dexter, MD, Rosemary Klein, MA, and Adrian Park, MD

Though in its infancy, the discipline of surgical ergonomics is increasingly valued. Still, little has been written regarding this field's tasks, models, and measurement systems. These 3 critical experimental components are crucial in objectively and accurately assessing joint and postural control as exhibited by expert laparoscopic surgeons. Such assessments will establish characteristic patterns important for surgical training. In addition, risk factors associated with both minimally invasive surgical instruments and the operating room environment can be identified and minimized. Our review focuses on evidence-based experimental ergonomic studies undertaken in the field of laparoscopic surgery. Publications were located through PubMed and other database and library searches. This article describes tasks, models, and measurement systems and considers their specific applications and the types of data obtainable with the use of each. Advantages and limitations, especially those of measurement systems, are compared and discussed. Future trends and directions believed necessary for optimal investigation and results are also addressed.

Keywords: surgical ergonomics; methodology; measurement systems; laparoscopy; review

The rapid acceptance of laparoscopic surgery as a clinical alternative to traditional open surgery sometimes obscures the fact that minimally invasive surgery (MIS) is still relatively new. Newer still is our realization of the physical demands MIS can make on its practitioners, but those are precisely the issues that the very recent discipline of surgical ergonomics seeks to understand and address. Like all disciplines that start out at once self-contained yet also multidisciplinary, the field of MIS ergonomic research is fast becoming broadened through a variety of different approaches, so numerous that those outside the field as well as many inside the field may possess scant knowledge regarding the vast methodologic array of tasks, models, and assessments in use.

Ergonomics examines and seeks to minimize risk factors between human beings and the tasks and environments that occupy them. Its historical origins are traceable to two editions (1700 and 1713) of De Morbis Artificum Diatriba, or Diseases of Workers, in which the job hazards workers encountered were characterized by Bernardo Ramazzini, often termed occupational medicine's founder,[1] and also to the 1857 article, "An essay on ergonomy, or science of labour, based on the laws of natural science," by W.B. Jastrzebowski, who is credited with the first use of the term derived from coupling the Greek words for work ("ergon") and for natural law ("nomos").[2] Despite those and other early influences, the science of ergonomics is still regarded as a fairly recent discipline, one just shy of its 60th anniversary as a formal body of knowledge.[3]

The relative youthfulness of ergonomics belies the significance of the contributions it has made in many occupational fields, including the military, athletics, and medicine, in addition to other environments. Fatigue, stress, and equipment use as factors affecting task performance were of great interest to the military during World War II. Indeed, to study the strain that might be experienced by air personnel flying long-range missions, a simulated cockpit was

From the Department of Surgery, University of Maryland, Baltimore.

Address correspondence to: Adrian Park, MD, University of Maryland, Department of Surgery, 22 S Greene St, S4B14, Baltimore, MD 21201; e-mail: apark@smail.umaryland.edu.

built by Kenneth Craik, inaugural director of the Applied Psychology Research Unit (Cambridge).[4] Among the many sports benefiting from ergonomic applications is track and field. Proper shoe design and material composition can greatly reduce fatigue level and improve running performance.[5] "Dispersion of attentional resources"—a unique measure recently established within the anesthesia work domain for evaluating the workplace awkwardness that results from simultaneous task performance—is typical of knowledge ascertained through ergonomic clinical research.[6] Ergonomic theory, design, and applications play an important role in everyday life, as is evidenced by accumulating research on everything from the effect of computer keyboard designs on users with upper extremity musculoskeletal disorders[7] to the collection and analysis, using motion capture and pressure sensors, of posture parameters and sitting strategies to determine car seat design that is both comfortable and ergonomically sound.[8]

The study of ergonomics in the surgical arena has acquired increased importance with the advent and widespread acceptance of laparoscopic procedures. Overall, the case has been made that its advantages often make MIS the preferable alternative for patients. Specifically, it falls to surgical ergonomics to make the case through problem definition, research surveys and studies, and data acquisition and analysis that MIS is often a demanding alternative for its surgical practitioners. This phenomenon results primarily from the different nature of the minimally invasive surgical environment, which surfaces new issues for the surgeon whose access to information and ability to move are limited in particular ways.

Though literature reviews of ergonomics in minimally invasive surgery have been undertaken, they remain few in number. In explanation of a newly coined term—minimal-access surgery (MAS)-related surgeon morbidity syndromes—a lengthy review was undertaken covering a broad array of issues identified as problematic, such as instrument design, operative display systems, and access ports in addition to injury mechanisms resulting from procedural technique.[9] A limited review covered ergonomic laparoscopic research accomplished within 3 broad categories: physical, sensorial, and cognitive.[10] A more comprehensive review detailed the variety of MIS ergonomic studies on visualization, manipulation, posture, and workload (mental and physical), as well as on the operating environment overall.[11] The effects of visual and haptic perception, often reduced, on surgical performance are reviewed in an article that also investigates research on force.[12]

In this review we focus on the methodology specifically used within the MIS ergonomic discipline for collection and analysis of data. In doing so, we examine and discuss the tasks and models germane to studies within this field. Our intention is to reveal the depth and breadth of the research approaches used at this young stage of MIS ergonomics, particularly through review and discussion of tasks, models, and measurements, and to suggest how improvement and development of these essential factors will profoundly affect future research and outcomes.

Tasks and Models

Two crucial design components in surgical ergonomic research are tasks and models. As laparoscopy evolves into a well-accepted, established discipline, the tasks involved in MIS procedures have increasingly become more standardized, gaining acceptance from MIS professional and accrediting organizations.[13-17] The result is that these tasks, for example, Fundamentals of Laparoscopic Skills (FLS), are increasingly used in MIS ergonomic research. Additionally, tasks such as partial circle-cutting have been created specifically to facilitate surgical ergonomic research. The term "models" refers to physical forms—in MIS often animal, artificially made, or simulated—that are meant to provide a realistic approximation of operating room (OR) conditions and human anatomy.

Static Tasks

Static tasks involve no motion. An example of a static task is the opening and closing of an instrument against a spring-loaded clip at a set resistance[18,19] while the instrument is held in a fixed position.[20]

Simple Navigation Tasks

Although they do not simulate any particular operative procedure, these tasks involve joint movement and instrument manipulation, thus permitting measurement. Through a variety of navigational skill tasks, fairly simple outcome metrics, including time and error, can be derived. These types of tracking tasks include navigation around an electrified wire course[21] and simple touching of labeled points on a target.[22,23]

Computerized formats of these simple navigation tasks exist. For example, the Dundee Endoscopic Psychomotor Tester (DEPT) evaluates the user's ability to navigate a 5-mm probe with one hand through a series of holes on a target plate, culminating in touching a plate behind the target plate.[24,25] Its successor, the Advanced Dundee Endoscopic Psychomotor Tester (ADEPT), evaluates the subject using both hands simultaneously—using one of 2 standard laparoscopic instruments with one hand to execute the original task, while the other manipulates the target.[26] In addition to the metrics of execution time, number of errors, and successful task completion, ADEPT is capable of measuring flight trajectory by recording instrument positions in 3D space.

Manipulation Tasks

The most fundamental laparoscopic skills, such as object manipulation, suturing, and cutting, comprise this group of tasks, which require bimanual coordination. Most studies have used as their model some form of standard laparoscopic trainer box, as doing so makes research less reliant on proprietary, often expensive or inaccessible technology. For example, small object manipulation into a small aperture, instrument-to-instrument rope passing, and cable tying have been used as a representative series of tasks for assessing laparoscopic muscle activation.[27] Other groups testing similar skills have used tasks involving transfers, shape cutting, point touching, and needle handling.[28-30] Cutting partial circles has been used for assessing operative table height[31] and monitor height.[32] A more difficult task that required lifting and cutting of various threads from a foam board[33] was used to examine the effects of 2D- versus 3D-viewing technology on task performance.

The most commonly used manipulation tasks have been variations of suturing and tying. Such study tasks and models have ranged from tying of knots only,[34-40] to suture placement and tying on an artificial surface,[41-44] to suturing of porcine enterotomies.[45-50] As much as possible, objective metrics have been developed for these types of tasks. For example, Cuschieri and colleagues developed a knot quality score related to the knot breaking or slipping force and the strength of the suture material itself.[51] Suturing tasks with their specific marked targets are easily graded based on accuracy of suture placement and time to completion. Enterotomy closures have been graded based on leak pressure.

Traditional Training Boxes

The traditional training box comes in various forms, but its basic construct remains the same. It is a confined space roughly approximating the abdominal cavity into which ports may be placed to allow instrument access. Visualization is achieved through a camera, which can be a laparoscope or simple charge-coupled device (CCD) camera. A multitude of different tasks such as object transfer, circle cutting, and bowel suturing can be conducted within this environment.

Virtual Reality Models

Virtual reality (VR) simulators are becoming increasingly lifelike as they incorporate haptic feedback and actual laparoscopic instruments, among other elements. They also come advantageously incorporated with a sizable number of reproducible and standardized measures, as well as a wide range of possible exercises. The simulated tasks used so far have not been operative simulations per se but instead have been simulations of drills designed to develop laparoscopic skills. For example, a needle manipulation drill on a LapSim (Surgical Science, Gothenburg, Sweden) has recently been used to study the potential benefit of armrests.[52] MIST-VR (Medical Education Technologies, Inc., Sarasota, Florida), with its variety of simulated tasks, has been used to measure the effect of cognitive distraction[53] and physical workload,[54] and the clipping and cutting tasks simulated by the Xitact 500 LS (Xitact SA, Morges, Switzerland) have been used to assess instrument handle types.[55] Despite the possibilities and potentials offered, no single simulator has gained widespread use in ergonomic analysis. In addition, to be proposed as optimal models for use in ergonomic studies, VR simulators must improve overall in terms of the still generally unrealistic quality of their replicated visual and haptic feedback, variables that may result in subjects making movements that would not be made in real or more realistic circumstances.

Intraoperative Model

As an ergonomic research environment in which to conduct task and movement research, the OR is completely realistic and completely validated. The drawbacks to use of the OR for such research, however, are considerable. Difficulties in getting ergonomic

equipment into the OR environment have barely been overcome. Two formidable issues encountered are sterilization and cumbersome wiring. Another aspect that must be considered is that such research within the OR would have to be properly limited to the expert as the subject, thus excluding comparative data on trainees or those with less experience. These obstacles have not, however, prevented all intraoperative ergonomic research. Surgeon posture,[56] camera operator movement,[57] and hand movements[58] have all been recorded during laparoscopic cholecystectomies. Other ergonomic studies, unhindered by OR limitations, have investigated display system,[59] operative flow, and procedures.[60-62]

Discussion

Task and model complexity in MIS ergonomic studies vary depending on the goal of each experiment. The data measured from simple static or manipulation tasks are relatively easy to analyze and provide basic ergonomic or performance information. With data obtained from more complex tasks that provide more comprehensive information, such as body control strategies, the experiment must be very carefully designed and performed to minimize variability within and between subjects. For instance, laparoscopic suturing is a widely used task in surgical ergonomic research. Although this is an essential skill, the performance of which is unofficially acknowledged as the hallmark of a skilled laparoscopic surgeon, it seldom represents more than a small fraction of the time spent in a laparoscopic procedure. Maneuvers such as exposure, retraction, and dissection are far more frequently employed tasks requiring more extended time expenditure, yet they are not often the subject of ergonomic analysis. Tasks such as these must be deemed worthy of research consideration. Additionally, more models and tasks for surgical ergonomic studies should be standardized in the manner of the FLS.

Sophisticated measurement systems built from a variety of technologies determine what types of data are obtainable through evaluative use of tasks and models. Among the most widely used are motion analysis for capture of body movement patterns and electromyography (EMG) analysis for monitoring muscle activation, although force plate analysis for evaluation of postural stability is gaining in importance.

Motion Analysis

Human movement is the result of complex processes involving the brain, spinal cord, peripheral nerves, muscles, bones, and joints. Motion capture technology, which evolved continuously from basic photography to sophisticated computer-aided motion analysis systems, allows study of these complex processes through analysis of kinematic data that represent the relative movements of segments connected with rotating joints. Motion capture systems have been used in various fields of application, including gait analysis,[63,64] sports science and athletic training,[65,66] biomechanics and neuroscience research,[67-69] and film and animation production.[70,71] Surgical ergonomics in laparoscopy is a relatively new field employing motion analysis. A wide variety of movements, ranging from the maneuvering of surgical instruments to the upper and/or lower body movements of surgeons, are captured as kinematic data and analyzed to provide detailed information about what constitutes ergonomic safety for surgeons performing MIS tasks and procedures.

Technologies

Throughout the development of motion analysis systems, different technologies have been used for accurate, objective movement measurements. Having considered all motion capture technologies, including ultrasound tracking and electric/mechanical goniometers, we posit that the most representative technologies, as evidenced by research group use, are orientation, electromagnetic, optical, and video-based motion analysis systems (Table 1).

An orientation sensor system measures 3 angular movements—yaw, pitch, and roll. Electromagnetic tracking systems use Faraday's law to measure the orientation and location data of each sensor. With optical motion capture systems (Figure 1), light-emitting diode (LED) strobes surrounding a camera lens transmit light into a measurement space from within which retroreflective markers—placed to represent joint and segmental landmarks—reflect light back to the camera. Recent optical motion capture systems are capable of handling hundreds of markers and reconstructing body movements in real time. With these relatively new systems, movement at each joint can be calculated in 3 directions (flexion/extension,

Table 1. Motion Analysis Technologies in Surgical Ergonomics

Technology	User groups	Components
Orientation sensor	Berguer et al,[31] Smith et al,[32,54] Kondraske et al[73]	Solid-state 3-axis pitch, roll, and yaw sensor
Electromagnetic tracking	Huber et al,[33] Rasmus et al,[57] Dosis et al,[58,83] Ridgway et al,[75] Datta et al,[74] Mackay et al,[76] Bann et al,[77] Khan et al,[78] Moorthy et al,[79,81] Hernandez et al,[80] Munz et al,[82] Aggarwal et al,[84] Smith et al[85]	Electromagnetic field generator and receivers
Optical tracking	Emam et al,[35,36,38,47-49] Person et al,[56] van Veelen et al,[59] Lee et al,[72,88,89] Patil et al,[86] Gillette et al[87]	Computer-controlled video camera and retroreflective markers
Video analysis	van Veelen et al,[29,40] Matern et al,[30,90] Joice et al[45,62]	Video recording and observation

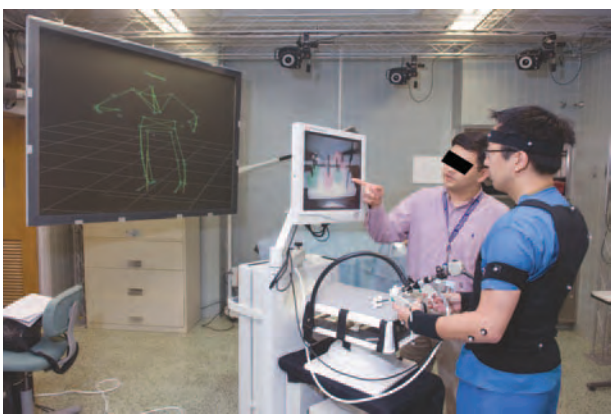


Figure 1. An optical motion analysis system with multiple cameras mounted on truss system tracks a set of reflective markers attached to the anatomical landmarks of a surgeon. The data is used to reconstruct the surgeon's body movement, which is then visualized as a stick diagram.

abduction/adduction, and internal/external rotation) and, for posture analysis specifically, the center of mass (CoM) can be calculated from combined kinematic and anthropometric data derived from subject measurement.[72]

The simplest system setup for video analysis involves a single camcorder or digital camera to monitor the subject's movement in 1 plane, for example, the sagittal plane. In surgical ergonomic studies, the endoscopic image has often been used to monitor performance accuracy and laparoscopic instrument movements. Video recordings captured by one or two cameras at different angles also provide approximate kinematic data analysis of surgeons' body movements.

Comparison of Motion Analysis Systems

By far the most economical motion analysis method in comparison with either electromagnetic or optical motion capture is the orientation sensor system. A significant limitation is that, although this system allows obtainment of absolute orientation information regarding each body segment to which an orientation sensor is attached, it does not record the locations of individual body segments; thus, full-joint kinematic analysis is very difficult.

The best overall accuracy and resolution is provided by an optical motion analysis system using digital cameras. With these systems, however, blocking problems are common, making optical motion tracking especially problematic for use in an actual OR, where obstacles located between markers and motion cameras may hinder or obscure detection of reflective markers. Another limitation associated with this system is that "ghost markers," or interfering noises, occur as a result of reflections from the metallic surfaces of surgical instruments or devices.

Electromagnetic tracking systems, although not plagued by the blocking issue, are bedeviled by data collection interference that, particularly in the OR theater, occurs as a result of other sources in the same field generating electromagnetic fields or as a result of metallic objects.[84] Moreover, their measurement volume is not spacious. Both issues result in measurement accuracy being significantly decreased when a sensor is not within transmitter range.

Video analysis is the simplest of such systems.[45,62] Although it has been used in ergonomic studies for accuracy assessment of task completion and movement tracking of laparoscopic instruments, the outcomes obtained by such video analysis are considered very subjective and qualitative. Thus, this system should not be used in a stand-alone manner, but only in conjunction with a quantitative motion capture system so that the two sets of data can be time synchronized for accurate assessment of body or surgical instrument motion.[35,58,83,86]

Table 2. Applications in Surgical Ergonomics Using Kinematic Data from Motion Analysis Systems

Assessments
Physical workload and fatigue[31,54]
Effect of operating room components:
Display systems[32,33,59]
Instrument handle design[38,40,44,49]
Instrument, task and scope location/orientation[36,47,48]
Operating table height[29,31]
Skill, dexterity, efficiency[35,44,73-80,82,85,88]
Others[45,62]

Motion Data Analysis

Kinematic data, the measurement outcome of motion analysis systems, are helpful in addressing a variety of surgical ergonomic issues. Table 2 shows important implications of motion analysis data. Berguer, Smith, and colleague used motion analysis in addition to EMG for measurement of upper arm elevation level to calculate physical effort/workload.[31,54] The effects on performance and body movements of changing variable elements in a variety of OR components—including display, instruments, and operating table—have been investigated. Attempts to determine what constitutes an optimal laparoscopic display system have been marked by obtaining and analyzing data, predominantly objective but also subjective (eg, self-reports), on head rotation, instrument tracking, and task accuracy in relation to a considerable array of display types, heights, and locations.[32,33,59] One study defined optimal table height as necessary for surgeons to achieve optimal task performance with minimal shoulder discomfort and workload. It suggested that the optimal table height could be easily found by having the instrument handle as held in the surgeon's hand and the surgeon's elbow at the same level.[31] Another study found optimal operating surface height to be between a factor 0.7 or 0.8 of elbow height.[29] Several studies have investigated how laparoscopic instrument handles, believed to be a primary cause contributing to unsafe ergonomics, can be redesigned to improve surgical performance and ergonomics.[38,40,44,49]

Cuschieri and colleagues have characterized joint movement in maximum movement, minimum movement, and range of motion (ROM), in addition to angular velocity. They have used these kinematic measures in their investigations on the effects of instrument, endoscope, and surgical task location and orientation. In doing so they have found appropriate intracorporeal and extracorporeal length ratios and alternative endoscope angles and been able to postulate on what constitutes proper orientation of task targets.[36,47,48]

Motion analysis data has also been used to describe skill and dexterity levels. Joint kinematic analysis can characterize surgical movement patterns used by expert surgeons,[35,44,88] compare different surgical techniques,[75,78] and assist in defining learning curves.[85] Darzi and colleagues defined motion efficiency by studying what was required for the hand to accomplish its objectives, specifically in terms of number of movements, length of travel path, and measurements of performance time. Those studies resulted in motion efficiency being defined as the least number of hand movements employed, a definition that has also been said to characterize the best kinematics and to provide the best quantifying measurement of surgical dexterity.[74-77]

Motion analysis additionally has been used in robot-assisted surgery and robotic camera control to assess the efficiency of these technologies.[73,79,80,82] In other studies, assessment of motion analysis system data has been applied to accuracy of surgical outcomes and to instrument tracking.[45,62]

Discussion

New developments and improvements in hardware, software, and research variables will extend the application of motion analysis in surgical ergonomics. Advancements in computers, digital video equipment, and digital technology allow multiple camcorders to be synchronized during motion capture, making possible 3D motion analysis. Although these advanced systems are capable of providing kinematic analysis in addition to digital imaging, they have not yet been used in surgical ergonomic research. The accuracy of research will be greatly improved by technologies in magnetic field generation and sensor detection that are being developed and incorporated to overcome noise interference and signal attenuation.

As motion analysis becomes a more common methodology in surgical ergonomics research, the demand will be that such systems provide more sophisticated data that will allow more refined, comprehensive assessment of issues ranging from fatigue to joint control strategies. Currently, traditional research variables, such as ROM, which characterizes the absolute difference between two extremes and is widely used to explain joint movement range, are too limited

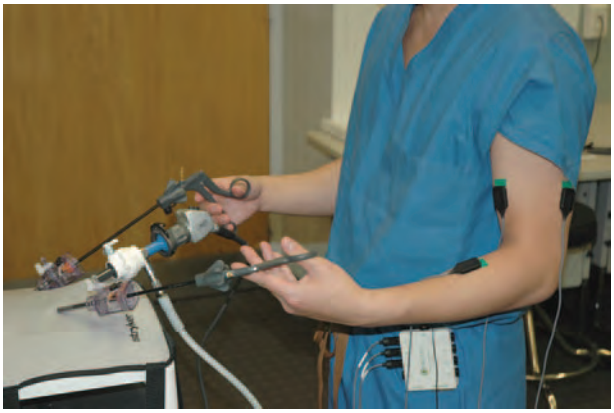


Figure 2. For EMG measurement, surface electrodes with built-in amplifiers are attached to target muscles.

to offer much information. One limitation of ROM, for instance, is that even a few extreme outliers may mislead and result in inaccurate interpretation.

Many kinematic variables used in surgical ergonomic studies, such as mean joint angle (MJA) or maximum angular velocity (MaxV), are calculated over a certain period of performance. Although these variables have been successfully used in surgical ergonomic analysis, the single values they provide cannot be used to explain details such as how or when joint angles change significantly or whether a specific pattern of joint movement exists. For more useful, comprehensive characterization of surgical joint movements, more research variables and analysis approaches will have to be developed to fill the gaps that traditional variables cannot explain. For example, Lee et al[88] used MJA and mean joint movement amplitude (MJMA), calculated from MJA and standard deviation, to study the joint movement range exhibited by very experienced and less experienced surgeons performing a pegboard transfer task. In another study, a pegboard transfer task was partitioned by subfunctions into several subtasks, and joint kinematics were investigated within those individual subtasks.[91]

As more sophisticated, comprehensive data sets emerge from motion analysis studies to give depth and definition to research, revisiting early surgical ergonomic study variables should become more commonplace. For example, motion efficiency has been defined by the number of hand movements used to complete a task. The relationship between motion efficiency and manual dexterity, however, is proving to be more complex than first thought (consider, for instance, that the instrument maneuverings by highly skilled surgeons might well strategically involve many hand movements) and certainly merits further consideration.

Electromyography

With laparoscopic procedures ever more prevalently used, MIS practitioners are reporting different patterns and types of fatigue and discomfort. The most common discomforts reported are located in the arms, neck, and upper back. EMG is one of the best objective tools for studying the mechanisms underlying the muscular discomfort or fatigue reported by laparoscopic surgeons, as EMG measures and evaluates the electrical activity of muscles in action and at rest. Ergonomists have long been acquainted with the many different ways EMG can be used to quantify physical workload and muscular discomfort.[92]

Technology

There are two types of electrodes used for EMG measurement: fine wire and surface. For evaluation with higher reliability of muscle fatigue and force, it has been proven that surface EMG is superior to fine-wire EMG.[93,94]

A fine-wire electrode made of thin and flexible metal is directly introduced into the target muscle to monitor the activation pattern of a single motor unit. Using a fine-wire electrode, it is possible to pick up an electrical signal without having attenuation caused by resistance of skin and tissues. The major limitations of the fine-wire electrode are its inability to measure the activity of a muscle as a whole and its invasive, often pain-causing nature.

Surface EMG is a simple, noninvasive way to evaluate the superficial muscles' activities (Figure 2). Though placement of the electrodes is relatively quick, painless, and sanitary, care must be taken to set them so as to minimize the effect of noise signals coming from more superficial or nearby muscles. Since skin and underlying tissues between the electrode and the target muscles serve as electrical resistance, the surface EMG provides only relative amplitude information, in contrast to the absolute values provided by fine-wire EMG. However, surface electrodes, owing to their considerable advantages, are dominant in surgical ergonomic studies for measurement of physical workload and upper body fatigue.

Electromyography Assessment Variables

EMG data provide fundamental information about amplitude (how much) and timing (when) of muscle activation. Analysis of EMG data allows characterization of the relationship between muscle activation and its outcomes (joint movement, force, or torque), as well as estimation of muscle fatigue level.

Currently, EMG data analysis is performed using one of two methods. The more basic measurement is a simple comparison of average or peak amplitudes of EMG signals over a certain period of time.[95] Such a basic comparison, although useful, cannot be used to compare activation amplitudes from different muscle groups used by an individual subject or the same muscle group used by a number of subjects. However, percentage of maximum voluntary contraction (%MVC) can be obtained by calculating the percentage ratio of measured EMG amplitude to a reference value, most commonly MVC. These normalized numbers can then be used for comparisons.

Frequency analysis is a third measurement used for muscle fatigue level assessment, a specific element of muscle activation. Muscle fatigue is perhaps the most significant ergonomic risk factor for surgeons, but use of amplitude and %MVC only indirectly infers its measurement.[96]

Electromyography Data Analysis

Amplitude analysis of EMG signals has been used in several studies in surgical ergonomics.[18,31,32,44,97] Berguer and colleagues compared EMG amplitudes to posit the optimal working angle for a laparoscopic instrument,[18] optimal ergonomic table height,[31] and proper monitor height.[32] Uchal et al[44] reported that no muscle activation differences occurred in a comparison of laparoscopic instrument handles, in-line vs pistol grip. Maithel et al[97] compared head-mounted and traditional video displays during simulated laparoscopic procedures and used EMG analysis to compare which display system caused more muscle fatigue.

Other studies used %MVC analysis.[19,21,28,29,41,98,99] Berguer and colleagues have shown that laparoscopic technique requires more physical effort than open surgical technique.[41,98] They also used the same analysis to compare different instrument grips.[19] Matern and colleagues have compared different monitor positions[99] and instrument handle designs.[21] Determination of optimal operating surface height has been another application.[29] Fatigue responses in several muscle groups have also been studied, with emphasis on effects of different monitor positions and levels of surgical experience.[28]

There are studies that have attempted to add other measurements to %MVC normalized EMG data. For example, Quick et al[27] have defined relative time of activation (RAT) as the percentage of time duration at which %MVC of each muscle group is 10% or higher. RATs were then calculated from individual muscle groups during different tasks, which proved useful to explain each muscle's activation timing during a specific task and to provide information for assessing muscle specific overuse. Jonsson has introduced the concept of amplitude probability distribution function (APDF) to evaluate the distribution of muscle contractions over a period of time. This approach identifies the percentage of time that muscle activity is below a preset proportion of MVC.[100,101] Jonsson also proposed the 10th percentile data that is commonly known as the static load level and has been used in surgical ergonomics to serve as a threshold in defining continuous work and minimal risk muscular load.[28]

Additionally, total physical workload as an assessment of fatigue has been obtained by calculating the time integral (that is, the area under the signal) of EMG over a specified period.[98] This approach, however, is appropriate only when each individual task is given a set time frame; it is not appropriate for evaluation of tasks unconstrained by time.

Discussion

Currently, the majority of EMG studies in surgical ergonomics investigate muscular workload during laparoscopic tasks through analysis of signal amplitude or %MVC. As muscular fatigue is a crucial variable in ergonomic risk analysis, it requires a more specific, accurate measurement tool. Frequency analysis, commonly considered the gold standard for studying muscular fatigue in biomechanic and neuroscience research,[102,103] has currently been used in only a few surgical ergonomic studies.[49,104] Frequency analysis should be used more intensively in our field.

Additionally, since the factor of muscle fatigue is so important, research tools must be developed that are capable of providing more detailed information in regard to issues that include: identification of muscular fatigue initiation, quantification of fatigue progress, and assessment of potential ergonomic risks of extreme fatigue. Also crucial is the need to

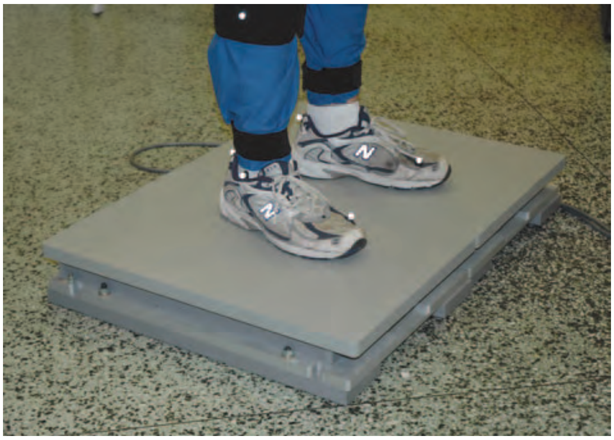


Figure 3. A surgeon performs a task with each foot on a single force plate. Data collected from each force plate is used for postural analysis.

determine risk levels within different muscle groups. This determination would be accomplished by establishing and considering those anatomical and geometric differences, such as type, size, length, strength, and frequency of use, that exist between and among muscles. If such detailed data could be generated and studied, ergonomics research could determine quantification and normalization through intermuscle comparisons, thereby acquiring vital knowledge about the vulnerability of individual muscles.

Force Plate Systems

Force plate—also called force platform—systems are a staple for scientific measurement of body posture within many motion analysis laboratories, yet they have seen limited use in surgical ergonomics (Figure 3).

Because the geometry of the human body consists of bigger and heavier masses at the upper body with supports at the lower body, posture during standing often has been portrayed as an inverted pendulum.[105] This is because even though quiet standing appears to be static, such posture actually relies on dynamics. The CoM is located at a short distance in front of the ankle joint, and therefore gravity alone can cause the body to topple forward. Our bodies are effectively engineered to keep us from falling face forward by two calf muscles—the soleus and the gastrocnemius—that actively compensate for gravity's effect and allow us to maintain proper static posture.[106]

Single or multiple force plate systems have been used to investigate both sway and balance control during quiet standing,[107,108] perturbed standing,[109] and functional standing or walking.[110,111]

Technology

A force plate provides valuable information about ground reaction force (GRF), which is equal in intensity and opposite in direction to that force exerted by the foot of the weight-bearing limb. The force plate usually has a rigid top plate made of a large piece of metal or glass. Force sensors at each of its 4 supporting corners produce an electric output proportional to the force applied to the upper surface and the contact point location.

Force sensors are either piezoelectric or strain gauge. Piezoelectric systems use quartz transducers to produce electricity that is proportional to the level of pressing. These systems require special cabling and charge amplifiers and generally have greater sensitivity and force measurement range than strain-gauge systems. However, piezoelectric systems may not be used for studies that involve prolonged standing, since the electrical charge attenuates over time. A strain-gauge system needs no special cables or amplifiers, and its electric signal does not decrease over time.

Use of a force platform makes obtainable the 3 components (vertical, lateral, and fore-aft) of force, the two coordinates of the center of pressure (CoP), and the rotational forces (moments) about the x, y, and z axis. The location of the CoP on a 2D surface has been widely used in postural analysis.

Force Plate Data

The importance of good posture for surgeons, as well as for other medical staff, is well documented.[112,113] Upper body postures, including head, arm, and trunk movements, have been analyzed in several studies[31-33,56,59,72,84] in attempts to define what constitutes optimal surgical posture, yet few studies in surgical ergonomics have used balancing information available from force plate systems alone.

Berguer et al[114] used force plate data to compare surgeons' postures during laparoscopic and open surgical procedures. Using a single force plate system and analyzing the locations and ROM of CoP in two directions (anterior-posterior and medial-lateral), the study concluded that during laparoscopic procedures, surgical posture was less dynamic, as was shown by significantly reduced ROM of CoP. Using primarily FLS tasks, Park and colleagues have studied changes in surgeons' postures during task performance.[72,87,89,115]

In the first study, they found that when CoP excursions were defined by outer boundaries, a significant increase in movements could be correlated to task difficulty.[87] Another study analyzed the postural control strategies used by surgeons of differing experience levels in the performance of 3 different FLS tasks. The more experienced surgeons used unique postural controls during each task, and their strategies differed from those used by less experienced surgeons.[89] To demonstrate that postural instability does not necessarily correlate to poor performance, another study used CoM as a measurement. A much-experienced surgeon with a wrist complication made compensatory arm movements to minimize wrist flexion and exhibited strategic movements that, although they appeared to signal postural instability, actually proved necessary to achieve successful task performance.[72] For a more systematic assessment, Lee and Park used principal component analysis (PCA) for construction of an ellipse that covered 95% of CoP excursions to characterize sway areas, directions, and shape.[115] In that same study, a new variable was termed and defined: postural stability demand (PSD), a calculation of the absolute distance between CoP and CoM. Through this rarely used measurement, it was discovered that during performance of each of 3 FLS tasks, less experienced surgeons exhibited high PSD, which implied higher postural instability.

Discussion

Maintaining good posture is absolutely necessary for top surgical performance. The balancing activity measured by force plate systems should be regarded as a necessity for postural analysis undertaken in surgical ergonomics.

Alone, the data obtained from force plates are useful. Ultimately, however, such data, if combined with motion analysis data, will allow calculations by which more detailed mechanical assessments may be acquired. Then, by applying biomechanical models (mathematical and/or analytical algorithms) to that more robust description, crucial information regarding the underlying etiology of the extreme muscular fatigue and physical complications often experienced by laparoscopic surgeons may be investigated.

Conclusions

We have intensively reviewed what we have identified as the primary research components—that is tasks, models, and measurement systems—that form the essential methodology of MIS ergonomics as this discipline moves through its initial maturation stages. In doing so, we have addressed elements including, but not limited to, background, technologies, applications, drawbacks, and advantages. In closing, we discuss limitations and possibilities we see inherent in these primary research components and proffer suggestions that if put into practice we believe will only augment the credible and useful findings already comprising surgical ergonomic research.

The quality and quantity of data collected are significantly dependent on tasks, models, and assessment systems. The research that has governed surgical ergonomics thus far has primarily addressed specifics through attempts to seek and obtain solutions that might be quickly applied to minimize the immediate effects of immediate problems. Current data collection, although useful, is limited in terms of what it can contribute to the formation of standard matrices.

For true effectiveness and validity, surgical ergonomic studies must expand in the near future to include more investigation and compilation of physical behavior characteristically engaged in by expert surgeons, including joint movement and postural control. Such research will promote extrapolation and description of characteristic patterns. Knowledge of such patterns is integral to the formation of standard matrices and will help to define what constitutes efficient, effective, and ergonomically safe physical behavior. Additionally, the creation and application of more objective training protocols promise to be important outcomes served by the development of such matrices.

Typically in a discipline's infancy, its studies address simple problems through simple constructs. For instance, it now appears that the range of acceptable movements or strategies employed by surgeons during task and procedure performance is far larger and more complex than envisioned in the early days of surgical ergonomic research. For specific illustration, early studies seemed to adequately explain that the fewer movements one used to complete a task (or the smaller the amount of overall motion involved in task execution) was indicative of what constitutes surgical maneuvering at its most effective and most efficient.

As our field becomes more established and sophisticated, however, it becomes more difficult to accept traditional concepts such as less or fewer are best. Discipline maturation requires that explanations

that gave meaning as research was conducted initially be subjected to scrutiny. Thus, the sanction accorded the concept of less or fewer movements is necessarily challenged by newer surgical ergonomic research that details the unique patterns of movements (sometimes many and sometimes few) creditable to each surgeon trying to successfully and competently achieve surgical goals while maintaining personal body comfort and stability.

Extended physical issues could well produce long-term effects resulting in surgeons abandoning what has long constituted their normal movement and replacing that with alternative compensatory movements in an effort to restore the originally enjoyed level of surgical movement. No matter how well-intentioned such compensatory movements might be, they may unintentionally cause secondary problems.

Secondary muscle strategies, for instance, employed as a result of compensatory movements, constitute unique patterns whose explorations should be undertaken in future surgical ergonomic studies. As a specific example, we offer that one muscle group experiencing extended fatigue might be expected to give over to another muscle group. The secondary group of muscles, taking over for the primary group, is not likely to have been trained, a situation that could well incur loss of performance accuracy.

Surgical ergonomics research is now addressing issues and seeking solutions that will result in levels of data acquisition that are still rare. Though our review indicates a paucity of data acquired within the OR, indications are that the challenges of safely and unobtrusively conducting ergonomic research in that environment will increasingly be undertaken. As VR models become more realistic and comprehensive, simulation in the laboratory environment will yield more sophisticated information. A broadening and deepening of the laparoscopic tasks studied (current research is too dependent on suturing) should also occur.

Measurements as well as assessment systems, often used across disciplines and in multiple combinations, but now only marginally considered within our discipline (as we have indicated has been the case with force plate/force platform systems), will with more prevalent use permit us to gather more detailed, robust data. And there is no doubt that the future surely holds in store for us the formation of large, cross-disciplinary databases in which the growing amount of surgical ergonomic research will be stored and referenced.

The ideal result of these efforts would be comprehensive characterization of surgical movement in relation to task or procedure. This result would then allow exact movements (with, presumably, their associated excellent clinical outcomes) to be patterned and then to be incorporated into the teaching of novice surgeons. Furthermore, and perhaps more urgently, such knowledge would facilitate better design not only of surgical instruments but also of the surgical workspace, which would lessen or eliminate the ergonomic ravages of MIS currently associated with laparoscopic surgical performance.

References

1. Franco G, Fusetti L. Bernadino Ramazzini's early observations of the link between musculoskeletal disorders and ergonomic factors. *Appl Ergon*. 2004;35:67-70.
2. Scheer SJ, Mital A. Ergonomics. *Arch Phys Med Rehabil*. 1997;78(Suppl 3):S36-S45.
3. Wilson JR. Fundamentals of ergonomic in theory and practice. *Appl Ergon*. 2000;31:557-567.
4. Craik K. Theory of the human operator in control systems. *Br J Psychol*. 1947;38:56-61.
5. Nigg BM, Stefanyshyn D, Cole G, Stergiou P, Miller J. The effect of material characteristics of shoe soles on muscle activation and energy aspects during running. *J Biomech*. 2003;36:569-575.
6. Seagull FJ, Ward R, Mills J, Goodrich C, Xiao Y. Measuring awkwardness of workplace layout: dispersion of attentional and psychomotor resources within the anesthesia workspace. *Proceedings of the Human Factors and Ergonomics Society 48th Annual Meeting*. 2004;48:1755-1758.
7. Tittiranonda P, Rempel D, Armstrong T, Burastero S. Effect of four computer keyboards in computer users with upper extremity musculoskeletal disorders. *Am J Ind Med*. 1999;35:647-661.
8. Andreoni G, Santambrogio GC, Rabuffetti M, Pedotti A. Method for the analysis of posture and interface pressure of car drivers. *Appl Ergon*. 2002;33:511-522.
9. Reyes DAG, Tang B, Cuschieri A. Minimal access surgery (MAS)-related surgeon morbidity syndromes. *Surg Endosc*. 2006;20:1-13.
10. Goossens RHM, van Veelen MA. Assessment of ergonomics in laparoscopic surgery. *Minim Invasive Ther Allied Technol*. 2001;10:175-179.
11. Berguer R. Surgery and ergonomics. *Arch Surg*. 1999; 134:1011-1016.
12. Xin H, Zelek JS, Carnahan H. Laparoscopic surgery, perceptual limitations and force: a review. On-line proceedings of the first Canadian student conference on biomedical computing. Available at: http://cscbc2006.cs.queensu.ca/assets/documents/Papers/paper144.pdf. Accessed May 30, 2007.

13. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. *Am J Surg*. 1998;175:482-487.
14. Derossis AM, Antoniuk M, Fried GM. Evaluation of laparoscopic skills: a 2-year follow-up during residency training. *Can J Surg*. 1999;42:293-296.
15. Feldman LS, Hagarty SE, Ghitulescu G, Stanbridge D, Fried GM. Relationship between objective assessment of technical skills and subjective in-training evaluations in surgical residents. *J Am Coll Surg*. 2004;198:105-110.
16. Peters JH, Fried GM, Swanstrom LL, et al; SAGES FLS Committee. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*. 2004;135:21-27.
17. Korndorffer JR Jr, Dunne JB, Sierra R, Stefanidis D, Touchard CL, Scott DJ. Simulator training for laparoscopic suturing using performance goals translates to the operating room. *J Am Coll Surg*. 2005;201:23-29.
18. Berguer R, Forkey DL, Smith WD. The effect of laparoscopic instrument working angle on surgeons' upper extremity workload. *Surg Endosc*. 2001;15:1027-1029.
19. Berguer R, Gerber S, Kilpatrick G, Remler M, Beckley D. A comparison of forearm and thumb muscle electromyographic responses to the use of laparoscopic instruments with either a finger grasp or palm grasp. *Ergonomics*. 1999;42:1634-1645.
20. Matern U, Giebmeyer C, Bergmann R, Waller P, Faist M. Ergonomic aspects of four different types of laparoscopic instrument handles with respect to elbow angle. *Surg Endosc*. 2002;16:1528-1532.
21. Matern U, Kuttler G, Giebmeyer C, Waller P, Faist M. Ergonomic aspects of five different types of laparoscopic instrument handles under dynamic conditions with respect to specific laparoscopic tasks: an electromyographic-based study. *Surg Endosc*. 2004;18:1231-1241.
22. Ahmed S, Hanna GB, Cuschieri A. Optimal angle between instrument shaft and handle for laparoscopic bowel suturing. *Arch Surg*. 2004;139:89-92.
23. Mishra RK, Hanna GB, Brown SI, Cuschieri A. Optimum shadow-casting illumination for endoscopic task performance. *Arch Surg*. 2004;139:889-892.
24. Hanna GB, Cuschieri A. Influence of the optical axis-to-target view angle on endoscopic task performance. *Surg Endosc*. 1999;13:371-375.
25. Hanna GB, Drew T, Clinch P, et al. A microprocessor-controlled psychomotor tester for minimal access surgery. *Surg Endosc*. 1996;10:965-969.
26. Hanna GB, Drew T, Clinch P, Hunter B, Cuschieri A. Computer-controlled endoscopic performance assessment system. *Surg Endosc*. 1998;12:997-1000.
27. Quick NE, Gillette JC, Shapiro R, Adrales GL, Gerlach D, Park AE. The effect of using laparoscopic instruments on muscle activation patterns during minimally invasive surgical training procedures. *Surg Endosc*. 2003;17:462-465.
28. Uhrich ML, Underwood RA, Standeven JW, Soper NJ, Engsberg JR. Assessment of fatigue, monitor placement, and surgical experience during simulated laparoscopic surgery. *Surg Endosc*. 2002;16:635-639.
29. van Veelen MA, Kazemier G, Koopman J, Goossens RHM, Meijer DW. Assessment of the ergonomically optimal operating surface height for laparoscopic surgery. *J Laparoendosc Adv Surg Tech*. 2002;12:47-52.
30. Matern U, Eichenlaub M, Waller P, Rückauer KD. MIS instruments: an experimental comparison of various ergonomic handles and their design. *Surg Endosc*. 1999;13:756-762.
31. Berguer R, Smith WD, Davis S. An ergonomic study of the optimum operating table height for laparoscopic surgery. *Surg Endosc*. 2002;16:416-421.
32. Smith WD, Berguer R, Nguyen NT. Monitor height affects surgeons' stress level and performance on minimally invasive surgery tasks. *Stud Health Technol Inform.* 2005;498-501.
33. Huber JW, Taffinder N, Russell RCG, Darzi A. The effects of different viewing conditions on performance in simulated minimal access surgery. *Ergonomics.* 2003;46:999-1016.
34. Crosthwaite G, Chung T, Dunkley P, Shimi S, Cuschieri A. Comparison of direct vision and electronic two- and three-dimensional display systems on surgical task efficiency in endoscopic surgery. *Br J Surg*. 1995;82:849-851.
35. Emam TA, Hanna GB, Kimber C, Cuschieri A. Difference between experts and trainees in the motion pattern of the dominant upper limb during intracorporeal endoscopic knotting. *Dig Surg*. 2000;17:120-125.
36. Emam TA, Hanna GB, Kimber C, Dunkley P, Cuschieri A. Effect of intracorporeal-extracorporeal instrument length ratio on endoscopic task performance and surgeon movements. *Arch Surg*. 2000;135:62-65.
37. Patil PV, Hanna GB, Frank TG, Cuschieri A. Effect of fixation of shoulder and elbow joint movement on the precision of laparoscopic instrument manipulations. *Surg Endosc*. 2005;19:366-368.
38. Emam TA, Frank TG, Hanna GB, Stockham G, Cuschieri A. Rocker handle for endoscopic needle drivers. *Surg Endosc.* 1999;13:658-661.
39. Hanna GB, Shimi SM, Cuschieri A. Task performance in endoscopic surgery is influenced by location of the image display. *Ann Surg.* 1998;227:481-484.
40. van Veelen MA, Meijer DW, Uijttewaal I, Goossens RHM, Snijder CJ, Kazemier G. Improvement of the laparoscopic needle holder based on new ergonomic guidelines. *Surg Endosc.* 2003;17:699-703.
41. Berguer R, Chen J, Smith WD. A comparison of the physical effort required for laparoscopic and open surgical techniques. *Arch Surg*. 2003;138:967-970.

42. Berguer R, Smith WD, Chung YH. Performing laparoscopic surgery is significantly more stressful for the surgeon than open surgery. *Surg Endosc*. 2001;15:1204-1207.
43. Moorthy K, Munz Y, Undre S, Darzi A. Objective evaluation of the effect of noise on the performance of a complex laparoscopic task. *Surgery*. 2004;136:25-30.
44. Uchal M, Brogger J, Rukas R, Karlsen B, Bergamaschi R. In-line vs pistol-grip handles in a laparoscopic simulator: a randomized controlled crossover trial. *Surg Endosc*. 2002;16:1771-1773.
45. Joice P, Hanna GB, Cuschieri A. Ergonomic evaluation of laparoscopic bowel suturing. *Am J Surg*. 1998;176: 373-378.
46. Brown SI, Frank TG, Shallaly E, Cuschieri A. Comparison of conventional and gaze-down imaging in laparoscopic task performance. *Surg Endosc*. 2003;17:586-590.
47. Emam TA, Hanna G, Cuschieri A. Comparison of orthodox vs off-optical axis endoscopic manipulations. *Surg Endosc*. 2002;16:401-405.
48. Emam TA, Hanna G, Cuschieri A. Ergonomic principles of task alignment, visual display, and direction of execution of laparoscopic bowel suturing. *Surg Endosc*. 2002; 16:267-271.
49. Emam TA, Frank TG, Hanna GB, Cuschieri A. Influence of handle design on the surgeon's upper limb movements, muscle recruitment, and fatigue during endoscopic suturing. *Surg Endosc*. 2001;15:667-672.
50. Ahmed S, Hanna GB, Cuschieri A. Optimal angle between instrument shaft and handle for laparoscopic bowel suturing. *Arch Surg*. 2004;139:89-92.
51. Hanna GB, Frank TG, Cuschieri A. Objective assessment of endoscopic knot quality. *Am J Surg*. 1997;174: 410-412.
52. Galleano R, Carter F, Brown S, Frank T, Cuschieri A. Can armrests improve comfort and task performance in laparoscopic surgery? *Ann Surg*. 2006;243:329-333.
53. Goodell KH, Cao CGL, Schwaitzberg SD. Effects of cognitive distraction on performance of laparoscopic surgical tasks. *J Laparoendosc Adv Surg Tech*. 2000; 16:94-98.
54. Smith WD, Berguer R. A simple virtual instrument to monitor surgeons' workload while they perform minimally invasive surgery tasks. *Stud Health Technol Inform*. 2004;363-369.
55. Matern U, Koneczny S, Tedeus M, Dietz K, Buess G. Ergonomic testing of two different types of handles via virtual reality simulation. *Surg Endosc*. 2005;19:1147-1150.
56. Person JG, Hodgson AJ, Nagy AG. Automated high-frequency posture sampling for ergonomic assessment of laparoscopic surgery. *Surg Endosc*. 2001;15:997-1003.
57. Rasmus M, Riener R, Reiter S, Schneider A, Feussner H. In vivo kinematic measurement during laparoscopic cholecystectomy. *Surg Endosc*. 2004;18:1649-1656.
58. Dosis A, Aggarwal R, Bello F, et al. Synchronized video and motion analysis for the assessment of procedures in the operating theater. *Arch Surg*. 2005;140:293-299.
59. van Veelen MA, Jakimowicz JJ, Goossens RHM, Meijer DW, Bussmann JBJ. Evaluation of the usability of two types of image display systems during laparoscopy. *Surg Endosc*. 2002;16:674-678.
60. Tang B, Hanna GB, Joice P, Cuschieri A. Identification and categorization of technical errors by observational clinical human reliability assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg*. 2004;139: 1215-1220.
61. Geryane MH, Hanna GB, Cuschieri A. Time-motion analysis of operation theater time use during laparoscopic cholecystectomy by surgical specialist residents. *Surg Endosc*. 2004;18:1597-1600.
62. Joice P, Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery—a human reliability analysis. *Appl Ergon*. 1998;29:409-414.
63. Lee G, Pollo FE. Technology overview: the gait analysis laboratory. *J Clin Eng*. 2001;26:129-135.
64. Giannini S, Catani F, Benedetti MG, Leardini A. *Gait Analysis: Methodologies and Clinical Applications*. Amsterdam, Netherlands: IOS Press; 1994.
65. Simon GSC, Andrew JR. A three-dimensional examination of the planar nature of the golf swing. *J Sports Sci*. 2005; 23:227-234.
66. Prassas S, Kwon YH, Sands WA. Biomechanical research in artistic gymnastics: a review. *Sports Biomech*. 2006; 5:261-291.
67. Kerrigan DC, Lelas JL, Goggins J, Merriman GJ, Kaplan RJ, Felson DT. Effectiveness of a lateral-wedge insole on knee varus torque in patients with knee osteoarthritis. *Arch Phys Med Rehabil*. 2002;82:889-893.
68. Siegler S, Liu W. Inverse dynamics in human locomotion. In: Allard P, Cappozzo A, Lundberg A, Vaughan CL, eds. *Three-Dimensional Analysis of Human Locomotion*. New York, NY: John Wiley & Sons; 1997.
69. Lee G, Fradet L, Ketcham CJ, Dounskaia N. Efficient control of arm movements in advanced age. *Exp Br Res*. 2006;177;78-94.
70. Menache A. *Understanding Motion Capture for Computer Animation and Video Games*. San Francisco, CA: Morgan Kaufmann Publishers; 1999.
71. Katherine P, Christoph B. Motion capture assisted animation: texturing and synthesis. Computer Graphics. Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques, San Antonio, TX, 23-26 July 2002. New York, NY: ACM Press; 2002.
72. Lee G, Kavic SM, George IM, Park AE. Postural instability does not necessarily correlate to poor performance: case in point. *Surg Endosc*. 2006;21:471-474.
73. Kondraske GV, Hamilton EC, Scott DJ, et al. Surgeon workload and motion efficiency with robot and human laparoscopic camera control. *Surg Endosc*. 2002;16: 1523-1527.
74. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessment. *Am J Surg*. 2002;184:70-73.

75. Ridgway PF, Ziprin P, Datta VK, et al. Laboratory-based validation of a novel suture technique for wound closure. *Ann Plast Surg*. 2002;49:291-296.
76. Mackay S, Datta V, Mandalia M, Bassett P, Darzi A. Electromagnetic motion analysis in the assessment of surgical skill: relationship between time and movement. *ANZ J Surg*. 2002;72:632-634.
77. Bann SD, Khan MS, Darzi A. Measurement of surgical dexterity using motion analysis of simple bench tasks. *World J Surg*. 2003;27:390-394.
78. Khan MS, Bann SD, Darzi A, Butler PEM. Use of suturing as a measure of technical competence. *Ann Plast Surg*. 2003;50:304-309.
79. Moorthy K, Munz Y, Dosis A, et al. Dexterity enhancement with robotic surgery. *Surg Endosc*. 2004;18:790-795.
80. Hernandez JD, Bann SD, Munz Y, et al. Qualitative and quantitative analysis of the learning curve of a simulated surgical task on the da Vinci system. *Surg Endosc*. 2004; 18:372-378.
81. Moorthy K, Munz Y, Undre S, Darzi A. Objective evaluation of the effect of noise on the performance of a complex laparoscopic task. *Surgery*. 2004;136:25-30.
82. Munz Y, Moothy K, Dosis A, et al. The benefits of stereoscopic vision in robotic-assisted performance on bench models. *Surg Endosc*. 2004;18:611-616.
83. Dosis A, Bello F, Moorthy K, Munz Y, Gillies D, Darzi A. Real-time synchronization of kinematic and video data for the comprehensive assessment of surgical skills. *Stud Health Technol Inform*. 2004:82-88.
84. Aggarwal R, Dosis A, Bello F, Darzi A. Motion tracking systems for assessment of surgical skill [letter to editor]. *Surg Endosc*. 2007;21:339.
85. Smith SGT, Torkington J, Brown TJ, Taffinder NJ, Darzi A. Motion analysis. *Surg Endosc*. 2002;16:640-645.
86. Patil PV, Hanna GB, Cuschieri A. Effect of the angle between the optical axis of the endoscope and the instruments' plane on monitor image and surgical performance. *Surg Endosc*. 2004;18:111-114.
87. Gillette JC, Quick NE, Adrales GL, Shapiro R, Park AE. Changes in posture mechanics associated with different types of minimally invasive surgical training exercises. *Surg Endosc*. 2003;17:259-263.
88. Lee G, Weiner M, Kavic SM, George IM, Park AE. Joint kinematics vary with performance skills during laparoscopic exercise [fundamentals of laparoscopic surgery (FLS) task 1]. *Gastroenterology*. 2006;130(Suppl 2):A911.
89. Lee G, Weiner M, Kavic SM, George IM, Park AE. Pilot study—correlation between postural stability and performance time during fundamentals of laparoscopic surgery (FLS) tasks. *Br J Surg*. 2006;93(Suppl 1):206.
90. Matern U, Kuttler G, Giebmeyer C, Waller P, Faist M. Ergonomic aspects of five different types of laparoscopic instrument handles under dynamic conditions with respect to specific laparoscopic tasks: an electromyographic-based study. *Surg Endosc*. 2004;18;1231-1241.
91. Lee G, Dexter DJ, Lee TH, Park AE. Subtask analysis of joint angles is key to characterizing surgical movement. *Gastroenterology*. 2007;132(Suppl 1):A894.
92. Koneczny S, Matern U. Instruments for the evaluation of ergonomics in surgery. *Min Invas Ther & Allied Technol*. 2004;13:167-177.
93. Gopher D, Donchin E. Workload—an examination of the concept. In: Boff KR, Kaufman L, Thomas J, eds. *The Handbook of Perception and Human Performance, Vol 2: Cognitive Processes and Performance*. New York, NY: Wiley; 1986.
94. Lofland KR, Mumby PB, Cassisi JE, Palumbo NL, Camic PM. Assessment of lumbar EMG during static and dynamic activity in pain-free normals: implications for muscle scanning protocols. *Biofeedback Self Regul*. 1995;20:3-18.
95. Hagg G, Luttmann A, Jager M. Methodologies for evaluating electromyographic field data in ergonomics. *J Electromyogr Kinesiol*. 2000;10:301-312.
96. Newton RU, Murphy AJ, Humphries BJ, Wilson GJ, Kraemer WJ, Hakkinen K. Influence of load and stretch shortening cycle on the kinematics, kinetics and muscle activation that occurs during explosive upper-body movements. *Eur J Appl Physiol*. 1997;75:333-342.
97. Maithel SK, Villegas L, Stylopoulos N, Dawson S, Jones DB. Simulated laparoscopy using a head-mounted display vs traditional video monitor: an assessment of performance and muscle fatigue. *Surg Endosc*. 2005;19: 406-411.
98. Berguer R, Forkey DL, Smith WD. Ergonomic problems associated with laparoscopic surgery. *Surg Endosc*. 1999;13:466-468.
99. Matern U, Faist M, Kehl K, Giebmeyer C, Buess G. Monitor position in laparoscopic surgery. *Surg Endosc*. 2005;19:436-440.
100. Ankrum DR. On the confusion between static load level and static task. *Appl Ergon*. 2000;31:545-546.
101. Jonsson B. Quantitative electromyographic evaluation of muscular load during work. *Scand J Rehabil Med Suppl*. 1978;6:69-74.
102. Lyons MF, Rouse ME, Baxendale RH. Fatigue and EMG changes in the masseter and temporalis muscles during sustained contractions. *J Oral Rehabil*. 1993; 20:321-331.
103. Kupa EJ, Roy SH, Kandarian SC, De Luca CJ. Effects of muscle fiber type and size on EMG median frequency and conduction velocity. *J Appl Physiol*. 1995;79: 23-32.
104. Judkins TN, Oleynikov D, Narazaki K, Stergiou N. Robotic surgery and training: electromyographic correlates of robotic laparoscopic training. *Surg Endosc*. 2006;20:824-829.
105. Gage WH, Winter DA, Frank JS, Adkin AL. Kinematic and kinetic validity of the inverted pendulum model in quiet standing. *Gait Posture*. 2004;19:124-132.

106. Loram ID, Maganaris CN, Lakie M. Paradoxical muscle movement in human standing. *J Physiol*. 2004;556.3: 683-689.
107. Masani K, Popovic MR, Nakazawa K, Kouzaki M, Nozaki D. Importance of body sway velocity information in controlling ankle extensor activities during quiet stance. *J Neurophysiol*. 2003;90:3774-3782.
108. Winter DA, Patla AE, Prince F, Ishac M, Gielo-Perczak K. Stiffness control of balance in quiet standing. *J Neurophysiol*. 1998;80:1211-1221.
109. Rietdyk S, Patla AE, Winter DA, Ishac MG, Little CE. Balance recovery from medio-lateral perturbations of the upper body during standing. *J Biomech*. 1999;32: 1149-1158.
110. Halliday SE, Winter DA, Frank JS, Patla AE, Prince F. The initiation of gait in young, elderly, and Parkinson's disease subjects. *Gait Posture*. 1998;8:8-14.
111. Karst GM, Venema DM, Roehr TG, Tyler AE. Center of pressure measures during standing tasks in minimally impaired persons with multiple sclerosis. *J Neurol Phys Ther*. 2005;29:170-180.
112. Kant IJ, de Jong LC, van Rijssen-Moll M, Borm PJ. A survey of static and dynamic work postures of operating room staff. *Int Arch Occup Environ Health*. 1992;63:423-428.
113. van Veelen MA, Nederlof EAL, Goossens RHM, Schot CJ, Jakimowicz JJ. Ergonomic problems encountered by the medical team related to products used for minimally invasive surgery. *Surg Endosc*. 2003;17:1077-1081.
114. Berguer R, Rab GT, Abu-Ghaida H, Alarcon A, Chung J. A comparison of surgeons' posture during laparoscopic and open surgical procedures. *Surg Endosc*. 1997;11:139-142.
115. Lee G, Park AE. Development of a novel tool to more precisely analyze postural stability of laparoscopic surgeons. *Surg Endosc*. 2007;21(Suppl 1):S317.

Surg Endosc (2007) 21: 471–474
DOI: 10.1007/s00464-006-9144-1

SURGICAL ENDOSCOPY
and Other Interventional Techniques

Postural instability does not necessarily correlate to poor performance: case in point

Gyusung Lee, Stephen M. Kavic, Ivan M. George, Adrian E. Park

Division of General Surgery, Department of Surgery, School of Medicine, University of Maryland, Baltimore, MD, USA

Received: 7 August 2006/Accepted: 22 September 2006/Online publication: 8 February 2007

Abstract
Background: It is very important for surgeons who perform minimally invasive surgery (MIS) to maintain proper postural stability, which kinematic research can determine. Previous studies in surgical ergonomics have shown that postural stability is correlated to instrument type, task difficulty, and skill level. What should also be considered is that surgeons may strategically change stance or joint movement to achieve better surgical outcomes while potentially subjecting themselves to greater risk. Background information about subjects, e.g., joint impairment, should be considered an important surgical ergonomic element. Such information can lead to more realistic and accurate conclusions about postural stability and joint kinematics.
Methods: A highly experienced and skilled right-handed surgeon developing carpal tunnel syndrome in both wrists was recruited into a small (6 subjects) performance study of pegboard transfer and circle-cutting tasks from the Fundamentals of Laparoscopic Surgery (FLS) skill set. Joint kinematics and postural data were collected using two associated force plates and a motion capture system of 12 digital, high-resolution, high-speed, infrared cameras.
Results: Each task was completed in less than 90 s. In pegboard transfer, the subject increased shoulder abduction angle to align his hand and forearm and minimize wrist flexion. When circle-cutting required excessive wrist flexion, the subject maintained his lower body position and stance while twisting his torso, a strategy that appeared to stabilize tangential direction related to cutting while maintaining a fixed orientation of forearm, wrist, and hand. In another circle-cutting trial, the subject changed his stance primarily by shifting foot position as necessary to obtain better scissor approach angles. These compensatory, strategic movements caused an increase in overall postural sway but did not represent postural instability.
Conclusion: This case study indicated that poor joint kinematics or postural stability does not necessarily correlate to poor performance. Instead, they may indicate positive compensatory or strategic movements.

Key words: Laparoscopy — Ergonomics — Kinematics — Posture — Fundamentals of Laparoscopic Surgery (FLS) — Motion analysis

Correspondence to: Adrian E. Park

Minimally invasive surgery (MIS) using laparoscopic instruments is attracting much attention because it promises patients less trauma, reduced infection risk, and quicker recovery than traditional surgery. However, laparoscopic surgery can be ergonomically unfavorable for surgeons manipulating long-shaft instruments with poor force transmission, working in limited access granted through ports (trocars), and viewing a 2D display while operating in a 3D work space [11, 12].

Maintaining proper postural stability is very important for surgeons performing MIS. Although a majority of studies in surgical ergonomics have examined performance time, accuracy, joint kinematics and/or electromyography (EMG) [1, 2, 14], and instrument tip movement [15], not many have included postural stability analysis. However, with the advent of laparoscopic surgery and its increased use, postural stability has become a crucial element for ergonomic research. Force plates, which were commonly used in biomechanical studies to collect postural data during quiet standing [7], perturbed standing [13], and functional standing or walking [3, 4], were appropriated for use in surgical ergonomic studies. In one study, Gillette et al. [2] showed that postural stability was correlated to instrument type and task difficulty. The pattern of force plate parameters varied with different graspers, and task difficulty was correlated to center of pressure (COP) excursion. Our recent study involving MIS surgeons with different levels of surgical experience showed a correlation between postural stability and performance time [5].

Table 1. Performance time and COM sway in M-L and A-P directions during pegboard transfer and circle-cutting tasks

	Performance time (s)	COM sway in M-L (mm)	COM sway in A-P (mm)
Pegboard transfer			
Subject A	55	28.5	31.53
Subject B	65.6	27.1	19.8
Subject C	80.4	15.0	20.4
Subject D	105.1	16.9	24.0
Circle-cutting			
Subject A, trial 1	58.3	110.7	76.4
Subject A, trial 2	87.3	98.6	37.9
Subject B	89.4	28.6	24.9
Subject C	76.8	60.5	33.1
Subject D	72.8	34.0	27.5

Subject A was the expert surgeon with carpal tunnel syndrome. Subjects B, C, and D were relatively experienced participants

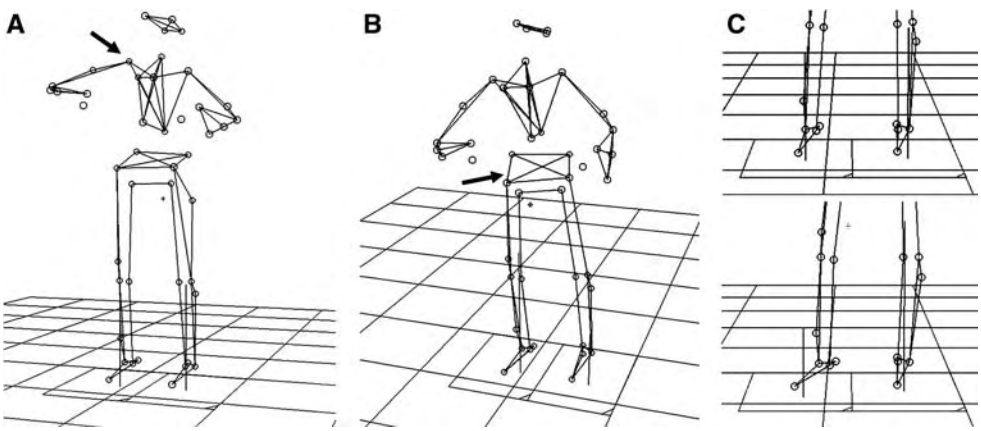


Fig. 1. Reconstructed body segments. Vertical lines at the feet represent ground reaction force vectors. **A** Pegboard transfer: the right shoulder abduction increased. **B** Circle-cutting trial 1: the torso twisted. **C** Circle-cutting trial 2: the right foot rotated externally.

The human body can accomplish the same kinematic outcome by using any number of combinations of movement patterns. This is because all body segments involved in movement are connected through joints, form a kinetic chain, and exhibit structural redundancy in degrees of freedom [6]. Previous studies did not consider that surgeons may make changes in their stance and/or joint movements as strategy-related or compensatory movements. This study presents a case of strategic and compensatory movements that worsened postural stability but not surgical performance.

Materials and methods

Six subjects with different levels of MIS experience were recruited from the Department of Surgery at the University of Maryland to complete pegboard transfer and circle-cutting tasks from the Fundamentals of Laparoscopic Surgery (FLS) skill set. These subjects volunteered for this study and signed an informed consent form. Endoscope images from a 0° scope were displayed on a standard CRT monitor positioned at eye level in front of the participants. Before each trial, participating surgeons watched instructional video clips that detailed the requirements of each task.

According to ViconPeak™ Plug-in-Gait™ (Lake Forest, CA) marker placement, thirty-nine 9.5-mm retroreflective sphere markers were attached to each participant. A state-of-the-art motion capture system also from ViconPeak™ (Lake Forest, CA), consisting of 12 high-speed, high-resolution, infrared, digital cameras, tracked the markers and reconstructed body segment movement in 3D space. The location data of these markers were sampled at 100 Hz. All tasks began when a verbal signal was given and were performed at a trainer box that rested on a height-adjustable table, allowing participants to maintain a right angle at the elbow while at rest. During each task, which was measured for performance time, participants stood with one foot on each AMTI (Watertown, MA) force plate, which sampled ground reaction forces and moment data at 1000 Hz. To quantify postural stability, the overall center of mass (COM) of each participant's entire body was calculated. We defined COM sway as the difference between the maximum and the minimum displacements of COM locations in medial-lateral (M-L) and anterior-posterior (A-P) directions. One of the participating surgeons, categorized as highly experienced, had developed carpal tunnel syndrome in both wrists. The data gathered on this subject are presented here.

Results

Task performance time and COM sway are summarized in Table 1. The data from subject A, the highly experienced surgeon with carpal tunnel syndrome, was compared with those of three other surgeons (subjects B, C, and D) who had been categorized as experienced. For the pegboard transfer task, subject A had the shortest performance time but greater COM sways in M-L and A-P directions. It was found that the subject increased the right shoulder abduction so that he could align his hand and forearm to minimize wrist flexion (Fig. 1A) and also increased the right shoulder flexion to place the lower arm into a more anterior position during the task. At the first trial of circle-cutting, subject A accomplished the task goal in less than 1 min. COM sways were very high in M-L and A-P directions. When excessive wrist flexion was required to cut a certain portion of the circle, subject A maintained his lower body position and stance

while twisting his torso. This strategy appeared to stabilize him in the tangential direction to the cutting while allowing him to maintain a fixed orientation of forearm, wrist, and hand (Fig. 1B). During the second trial of circle-cutting, subject A required more time than in the first trial but did not exceed the time required by subject B. COM sway in both directions was higher for subject A than for other subjects. During this trial, subject A changed his stance primarily by shifting foot position as necessary to obtain better approach angles for the scissors and his body weight was supported more by his right leg (Fig. 1C).

Further investigation will be required for a better understanding of the effect of each compensatory and strategy-related movement on postural stability. Such knowledge can contribute to a more accurate ergonomic definition of postural stability as related to those practicing laparoscopic surgery. Though this study used the difference between the maximum and the minimum COM displacements in two directions as its definition of overall COM sway, in future studies we plan to use COM sway patterns characterized by the outer contour of COM excursion. Using this variable will allow sway to be examined in all directions and kinematics to be correlated more accurately to postural stability. In addition, while many biomechanical studies use either COP, which is obtained from force plates and explains ankle activity to maintain stability, or COM, which is calculated from kinematics and represents the effectiveness of the ankle control, our use of these distinct yet closely related variables together promises a better understanding of the mechanisms involved in posture control [8, 9]. Our future studies will include the difference between COP and COM displacement in the hope that it will provide insights (e.g., the shorter the COP - COM difference, the higher the risk for postural instability) regarding postural control mechanisms [10].

In summary, during both the pegboard transfer and the circle-cutting tasks, the experienced surgeon with carpal tunnel syndrome in both wrists made compensatory and strategic movements. These movements caused an increase in overall postural sway, yet they did not necessarily represent postural instability.

Discussion

This case study showed that poor postural stability or joint kinematics does not necessarily correlate to poor performance but may instead be positive, that is, compensatory or strategic. Still to be addressed, however, is whether surgeons who may be strategically changing their stance or joint movement for the positive purpose of achieving better surgical outcomes may be subjecting themselves to negative kinematic risk. We propose expanded ergonomic research to investigate the effects of certain complications, e.g., carpal tunnel syndrome, neck pain, back pain, and osteoarthritis at the lower extremities as experienced by laparoscopic/endoscopic surgeons, especially those who have been practicing MIS for a considerable duration. By comparing the biomechanical data of MIS surgeons who possess such existing complication(s) with a normal control group, we can determine much about compensatory and/or strategic body movements as adaptations to physical conditions and can make complication-specific recommendations regarding ergonomic safety for MIS surgeons. Equally important will be to explore the mechanisms causing biomechanical secondary injury. Even though compensatory movement may help to relieve pain or discomfort caused by primary injury and strategic movement can position surgeons to best accomplish surgical goals, these movements may have an effect on other parts of the body in the same kinetic chain and initiate biomechanical secondary injury. This study has affirmed that background information about a subject's physical ailments, e.g., joint impairment, must be considered a vital element of surgical ergonomic studies. Correlations of such information are predicted to yield more accurate, specific, and useful conclusions about both postural stability and joint kinematics.

Acknowledgments. This study was supported by a grant from the U.S. Army Medical Research and Materiel Command (USAMRMC), and equipment was provided in part by U.S. Surgical Corp. (Norwalk, CT). The authors acknowledge the thoughtful assistance of Rosemary Klein in the editing of the manuscript.

References

1. Emam TA, Hanna G, Cuschieri A (2002) Comparison of orthodox vs off-optical axis endoscopic manipulations. Surg Endosc 16: 401–405
2. Gillette J, Quick N, Adrales G, Shapiro R, Park AE (2003) Changes in postural mechanics associated with different types of minimally invasive training devices. Surg Endosc 17: 259–263
3. Halliday SE, Winter DA, Frank JS, Patla AE, Prince F (1998) The initiation of gait in young, elderly, and Parkinson's Disease subjects. Gait Posture 8: 8–14
4. Karst GM, Venema DM, Roehr TG, Tyler AE (2005) Center of pressure measures during standing tasks in minimally impaired persons with multiple sclerosis. J Neurol Phys Ther 29(4): 170–180
5. Lee G, Kavic SM, George IM, Park AE (2006) Correlation between postural stability and performance time during fundamentals of laparoscopic surgery (FLS) tasks (poster). Annual conference of the association of surgeons of Great Britain and Ireland (ASGBI), Edinburgh, Scotland. Br J Surg 93(Suppl s1): 206
6. Lutz GE, Palmitier RA, An KN, Chao EYS (1993) Comparison of tibiofemoral joint force during open-kinetic chain and closed-kinetic chain exercises. J Bone Joint Surg Am 75(5): 732–739
7. Masani K, Popovic MR, Nakazawa K, Kouzaki M, Nozaki D (2003) Importance of body sway velocity information in controlling ankle extensor activities during quiet stance. J Neurophysiol 90: 3774–3782
8. Murray MP, Seireg A, Scholz RC (1967) Center of gravity, center of pressure, and supportive forces during human activities. J Appl Physiol 23: 831–838
9. Nault ML, Allard P, Hinse S, Le Blanc R, Caron O, Labelle H, Sadeghi H (2002) Relations between standing stability and body posture parameters in adolescent idiopathic scoliosis. Spine 27(17): 1911–1917
10. Panzer VP, Bandinelli S, Hallet M (1995) Biomechanical assessment of quiet standing and changes associated with aging. Arch Phys Med Rehabil 76: 151–157
11. Parkin M, Isabel L (1995) Ergonomics, engineering, and surgery of endosurgical dissection. J R Coll Surg Edinb 40: 120–132

12. Reyes DAG, Tang B, Cuschieri A (2006) Minimal access surgery (MAS)-related surgeon morbidity syndromes. Surg Endosc 20: 1–13
13. Rietdyk S, Patla AE, Winter DA, Ishac MG, Little CE (1999) Balance recovery from medio-lateral perturbations of the upper body during standing. J Biomech 32: 1149–1158
14. Smith SGT, Torkington J, Brown TJ, Taffinder NJ, Darzi A (2002) Motion analysis. Surg Endosc 16: 640–645
15. Zheng B, Verjee F, Lomax A, MaxKenzie CL (2005) Video analysis of endoscopic cutting task performed by one vs two operators. Surg Endosc 19: 1388–1395

Surg Endosc
DOI 10.1007/s00464-008-0141-4

Ergonomic risk associated with assisting in minimally invasive surgery

Gyusung Lee · Tommy Lee · David Dexter · Carlos Godinez · Nora Meenaghan · Robert Catania · Adrian Park

Received: 10 April 2008 / Accepted: 13 August 2008

Abstract
Background Given the physical risks associated with performing laparoscopic surgery, ergonomics to date has focused on the primary minimally invasive surgeon. Similar studies have not extended to other operating room staff. Simulation of the assistant's role as camera holder and retractor during a Nissen fundoplication allowed investigation of the ergonomic risks involved in these tasks.
Methods Seven subjects performed camera navigation and retraction tasks using a box trainer on an operating room table that simulated an adult patient in low lithotomy position. Each subject stood on force plates at the simulated patient's left side. A laparoscope was introduced through a port into the training box with four 2-cm circles as rear-panel targets located in relation to the assistant as distal superior, proximal superior, distal inferior, and proximal inferior target effects. The subjects held the camera with their left hand, pointing it at a target. The task was to match the target to a circle overlaid on the monitor. Simultaneously, a grasper in the right hand grasped and pulled a panel-attached band. A minute signal moved the subject to the next target. Each trial had three four-target repetitions (phase effect). The subjects performed two separate trials: one while holding the camera from the top and one while holding it from the bottom (grip effect). A $4 \times 3 \times 2$ (target $\times$ phase $\times$ grip) repeated-measures design provided statistics. Dividing the left force-plate vertical ground reaction forces (VGRF) by the total VGRF from both plates provided a weight-loading ratio (WLR).
Results The WLR significantly increased ($p < 0.005$) with proximal targets (2 by 80% and 4 by 79%). The WLR decreased 75%, 74%, and 71% over time. No difference existed between the grip strategies (grip effect, $p > 0.5$).
Conclusions A high-risk ergonomic situation is created by the assistant's left or caudal leg disproportionately bearing 70–80% of body weight over time. A distance increase between the camera head location and the camera holder increases ergonomic risk. The phase effect was interpreted as a compensatory rebalancing to reduce ergonomic risk. Ergonomic solutions minimizing ergonomic risks associated with laparoscopic assistance should be considered.

Keywords Camera assistant · Ergonomics · Force plate · Laparoscopic assistance · Postural analysis · Simulation

G. Lee · T. Lee · D. Dexter · C. Godinez · N. Meenaghan · R. Catania · A. Park (✉)
Division of General Surgery, Department of Surgery, School of Medicine, University of Maryland, 22 South Greene Street, Room S4B14, Baltimore, MD 21201, USA
e-mail: apark@smail.umaryland.edu

G. Lee
e-mail: glee@smail.umaryland.edu

T. Lee
e-mail: tlee@smail.umaryland.edu

D. Dexter
e-mail: ddexter@smail.umaryland.edu

C. Godinez
e-mail: cgodinez@smail.umaryland.edu

N. Meenaghan
e-mail: nmeenaghan@smail.umaryland.edu

R. Catania
e-mail: rcatania@smail.umaryland.edu

It is well known that maintaining correct posture is a very important ergonomic factor in minimizing physical risks associated with the performance of complex tasks [1–3].

Awkward working posture can cause increased stress for certain body parts, resulting in fatigue, musculoskeletal disorders, and nerve problems [4].

To maintain proper balance, the human body requires continuous active control. Because two-thirds of body mass ordinarily is in the top two-thirds of a person's height above ground level, the body has been described as an unstable balance system, and the standing posture often is referred to as an inverted pendulum [5]. For this reason, posture during quiet standing actually relies on dynamic control although the posture may appear to be static [6].

Many everyday tasks consist primarily of static posture while the upper body performs more dynamic motions (e.g. cashiers at registers, secretaries at computers, bus drivers). Ergonomic evaluations and assessments have been undertaken in various industrial workplaces to address the levels of postural stress/discomfort quantitatively for descriptions of optimal work postures [1, 7]. Increasing numbers of posture studies involving health care workers have been conducted, but their focus has been primarily on low back pain problems experienced by hospital nurses.

Ergonomic stress during dynamic tasks such as patient lifting and transferring is only part of a broader picture. Prolonged static posture has been associated also with ergonomic stress [8]. The physical stress associated with the fixed work posture of many surgeons and operating room staff can result in discomfort, fatigue, and musculoskeletal disorders.

Performing a laparoscopic surgical procedure places particularly high physical and cognitive demands on surgeons that differ substantively and dramatically from those experienced during open surgery. Knowledge of ergonomics related to primary laparoscopic surgeons has been well described in previous studies. Several minimally invasive surgery (MIS) components including long-shaft instruments, access ports, and endoscopic image display systems have been identified as contributing to ergonomically unfavorable postures assumed and maintained by laparoscopic surgeons during procedural performances [9–12].

Motion analysis and force plates have been among the tools used to examine surgeons' body movements, specifically the measurement and analysis of postural sway [13]. Upper body movements have been characterized by Person et al. [14], who demonstrated the feasibility of using an optoelectric measurement system for automated posture sampling in the study of surgical ergonomics.

Berguer et al. [15] compared surgeons' postures during laparoscopic and open surgical procedures by analyzing the locations and range of motion (ROM) of the center of pressure (COP) in two directions (anteroposterior and mediolateral). They concluded that during laparoscopic procedures, surgical posture is less dynamic, as shown by significantly reduced ROM of COP.

Gillette et al. [16] calculated the outer boundaries to quantify the amount of sway in COP excursions. They found that significant increases in movement could be correlated with task difficulty. Additionally, postural control strategies used by surgeons of differing experience levels have been analyzed during fundamentals of laparoscopic surgery task performance. These studies concluded that each task required unique postural control mechanisms and that a significant difference in sway control was evident among surgeons with different surgical skill levels [17, 18].

For a more systematic assessment of postural control, Lee and Park [19] characterized sway areas, directions, and shape by constructing an ellipse using principal component analysis that covered 95% of COP excursions. In that same study, postural stability demand (PSD) was calculated as the absolute distance between COP and the center of mass. This analysis showed that less experienced surgeons exhibited high PSD, which implied higher postural instability during performance of laparoscopic surgery task fundamentals.

Another postural study using center of mass sway analysis demonstrated that postural instability does not necessarily correlate with poor performance [20]. A highly experienced surgeon with a wrist condition made compensatory arm movements to minimize wrist flexion. This surgeon also exhibited strategic movements that although appearing to signal postural instability, actually proved to be necessary for achieving successful task performance.

The ergonomics associated with the operating surgeon's performance are understandably a priority. Although an active part of the entire operation, the laparoscopic assistant has not until currently been properly the subject of ergonomics studies.

Nissen fundoplication, a common procedure, involves both advanced and basic laparoscopic skills. The inexpensive, easily constructed box trainer we designed provides a simulation alternative to the animal models that to date have dominated studies of minimally invasive fundoplication task performance.

While performing a fundoplication, the surgeon stands centered between the patient's legs in alignment with the diaphragmatic hiatus. This position keeps the surgeon facing straight ahead in as neutral a posture as possible and one assumed to be ergonomically favorable. In contrast, the camera holder is positioned to accommodate the surgeon, which usually means standing to one side of the patient and thus not in optimal alignment with the working area. To accommodate, the camera holder is forced to rotate his or her upper body while simultaneously reaching across the operative field to hold the camera or retract tissues for the surgeon.

This study aimed to quantify ergonomic risks associated with the tasks performed during a fundoplication by the

camera assistant. The particular segment of the study reported involves postural balancing. We hypothesized that postural balancing is affected by the grip used to hold the camera and the location and relocation of the camera head in relation to different targets. We also investigated how fatigue influences postural balancing over time-framed (early, middle, late) phases of task performance.

Materials and methods

This institutional review board–approved study was performed in the Surgical Ergonomics Laboratory at the Maryland Advanced Simulation, Training, Research, and Innovation (MASTRI) Center at the University of Maryland. Seven right-handed subjects possessing different levels of MIS experience ranging from medical student to fellow from the Department of Surgery at the University of Maryland School of Medicine volunteered, signed an informed consent, and agreed to perform camera navigating and retracting tasks.

An adult patient in low lithotomy position was simulated by a training box on an operating room table. An extended arm board was attached to the table bearing the trainer box as a representation of the obstruction that would be caused by the patient's left leg. Additionally, to simulate the situation of the assistant working around the arm of the operating surgeon, we inserted two graspers through ports set on the right and left sides of the camera port and instructed each subject to work around these graspers. At the left side of the simulated patient, each subject stood on two force plates (Advanced Mechanical Technology Inc., Watertown, MA, USA), one leg on each (Figs. 1 and 2).

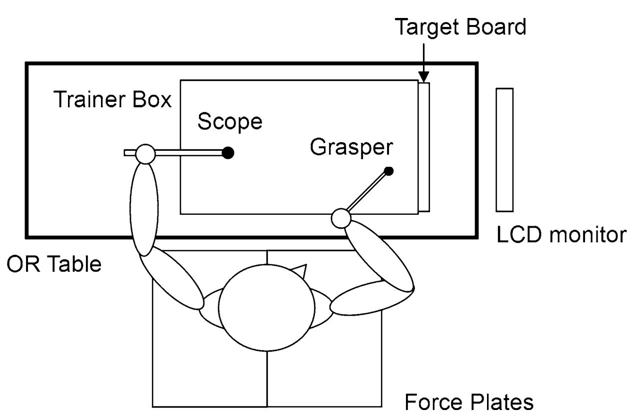


Fig. 1 Experimental setup for the simulation of the camera assistant's roles as a camera holder and retractor during a Nissen fundoplication. A trainer box placed on the operating room table simulates a patient in low lithotomy position. A board with targets, circles, and bands is attached to the side of the trainer box at the location of the simulated patient's head. While performing the tasks, the subject stands on two force plates

A 0° scope displayed endoscope images on a standard liquid crystal display (LCD) monitor positioned at eye level at the head of the bedside. A laparoscope was introduced into the training box, which also contained four 2-cm circles functioning as targets placed on the rear panel in the following relations to the assistant: distal superior, proximal superior, distal inferior, and proximal inferior (target effect). Four additional 2-cm circles approximately 10 cm apart were marked in a straight line between the superior and inferior targets. A rubber band was stapled to each of the outermost circles (A, C) for the purpose of retraction to the inner circles (B, D) (Fig. 3).

Each subject performed a 12-min trial that consisted of three 4-min phases (early-, middle-, and late-phase effect). In each phase of four dual-task modules, the subject performed camera-holding and -pointing tasks (moving from target 1 through target 4) together with grasping and retraction tasks (moving from A to B, C to D, A to B, and C to D). Each subject performed two trials, one while holding the camera from the top and one while holding it from the bottom (grip effect).

Specifically, the camera-holding and -pointing task required holding the camera with the left hand and pointing it at a target (Fig. 4). On the screen were two circles (diameters of 4.5 cm for the larger and 2.5 cm for the smaller) printed on a transparency attached to the monitor. The task for the subject was to maintain the target's accuracy constraints by confining it between the boundaries of both circles. The grasping and retraction task required holding a grasper in the right hand to grasp a rubber band at one location and retract it to another. Both tasks were always executed simultaneously.

Once both retraction and camera pointing were completed, the subject was asked to maintain accuracy constraints of both the grasping and camera-pointing tasks for a minute. While doing so, each subject was free to change or adjust his or her posture.

The amplitudes of the vertical ground reaction forces (VGRF) exerted by both the left and right legs onto each force plate were collected and recorded at 200 Hz. While each subject maintained static posture for each of the four tasks within each of the three phases, data permitting calculation of a weight-loading ratio (WLR) were obtained. To quantify the balancing taking place between the two legs, we derived the WLR by dividing the left force plate VGRF by the total VGRF from both plates.

$$WLR = \frac{\text{Left VGRF}}{\text{Left VGRF} + \text{Right VGRF}} \times 100$$

Thirty-nine 9.5-mm retroreflective sphere markers were attached to each participant according to Plug-in-Gait (ViconPeak, Lake Forest, CA, USA) marker placement designations. The markers were placed on bands around the

Fig. 2 Each subject while standing on two force plates performs the camera-pointing task with the left hand and the retraction task with a grasper in the right hand. Reflective markers and electromyographic electrodes are attached to the subject for motion and muscle activation analysis

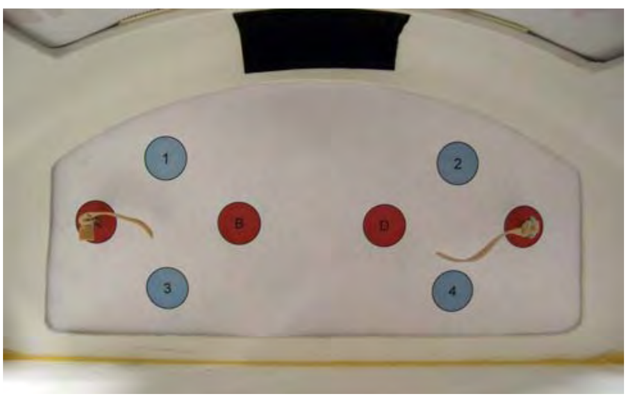


Fig. 3 Inside view of the trainer box showing the target board. Four circles numbered 1 to 4 are the targets for the camera-pointing task. Another four circles are labeled A to D. Circles A and D have rubber bands attached for the tissue retraction task

Fig. 4 Screen shot of the monitor showing a target confined between the two circles on the monitor

head, wrists, and knees; a custom-made vest and waist belt; bare arms; and the shoulder, thighs, and shanks of the participants' medical scrubs. Placement of elastic bands and a vest over the scrub suit of each participant permitted the markers to be attached securely to anatomic landmarks. A (ViconPeak, Lake Forest, CA, USA) motion capture system consisting of 12 high-speed, high-resolution, infrared digital cameras tracked the markers and reconstructed body segment movement in three-dimensional (3D) space. The location data of each marker were sampled at 100 Hz.

Statistical analysis

An overall $4 \times 3 \times 2$ (target $\times$ phase $\times$ grip) analysis of variance (ANOVA) with repeated measures was applied to characterize the extent to which each laparoscopic assistant's body weight was distributed by each leg. Then the main effects of these three factors (target, phase, grip) and their interactions were analyzed. The significance level was set at a p value of 0.05.

Results

The data on the two grip strategies collected separately showed no statistical difference between them in terms of weight balancing ($p > 0.5$). Therefore, the data were consolidated for further analysis using 4×3 (target $\times$ phase) repeated-measures ANOVA. The significant

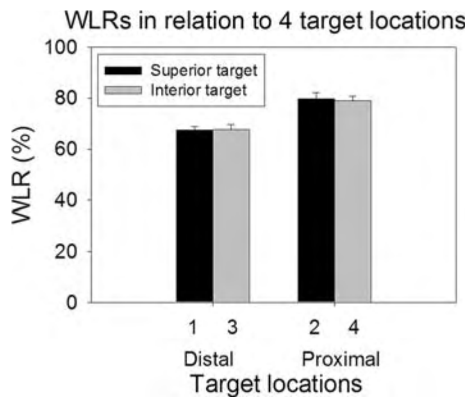


Fig. 5 Target effect as it relates to the weight-loading ratio (WLR), which significantly increased with proximal targets (i.e., 2, 4) by 11.8% compared with distal targets (i.e., 1, 3). No difference was found between targets on the same side (i.e., inferior, 3 and 4; superior, 1 and 2)

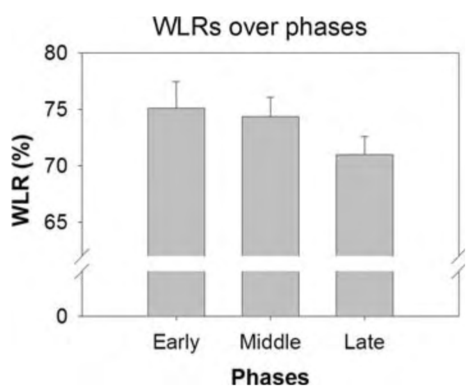


Fig. 6 Phase effect on weight-loading ratio (WLR), which decreased as each four-target phase of task performance was repeated

main result was that the WLR was found to be different among the four targets ($p < 0.005$), as shown in Fig. 5. Post hoc tests further showed that the WLR significantly increased with proximal targets (i.e., 2, 4) by 11.8% compared with distal targets (i.e., 1, 3). No difference was found between targets on the same side (inferior, 3 and 4; superior, 1 and 2). The findings also showed that WLR decreased as the four target camera navigation modules were repeated (phase effect, $p < 0.05$) (Fig. 6). The interaction effect between the target and phase factors was not significant ($p > 0.05$).

Discussion

This study investigated the ergonomic risks potentially experienced by laparoscopic assistants during a simulated laparoscopic Nissen fundoplication. Specifically, it analyzed the postural balancing that occurs as camera navigation and target retraction tasks are performed. Postural balancing analysis, achieved by what we termed WLR, using a force-plate system, demonstrated that the assistant's left leg disproportionately bore 70% to 80% of body weight over time, thus creating a high-risk ergonomic situation. It can reasonably be assumed that if the camera holder stood on the patient's right side, the data would be a mirror image of that presented in this discussion (i.e., more load on the right leg).

Additionally, after introduction of the camera through the one access port used in our study, we discovered an ergonomic risk traceable to the fulcrum effect. Specifically, when the camera was pointed toward the proximal target, the camera head actually moved toward the distal location. This is referred to as the target effect. The risk presented in this situation is that that the camera assistant must maintain continuous extension of the left arm while simultaneously leaning the entire body left.

When we considered the fatigue effect on postural balancing over phases, we found that WLR decreased as a phase was repeated (early, middle, late). This is referred to as the phase effect. We interpreted this WLR result as indicative of the camera assistant acting in a compensatory manner to combat the increased muscular fatigue that developed over time. Considering that this compensatory action resulted in a reduction of ergonomic stress at the joint of the left leg, we propose ergonomic solutions such as camera-handle attachments to minimize the time-accrued effects of the extended arm and unbalanced leaning posture. Another solution to minimize fatigue and maximize postural stability may be to train the camera assistant in a simulated situation, as done in our study, before actual operating room performance of a fundoplication, with specific instructions given on the need to rebalance weight as accurate camera pointing and retraction are repeatedly achieved.

To mitigate obstructions that potentially create positioning issues, compensatory action also is probable such as equipping the operating table with stirrups or having the assistant work around the arms of the operating surgeon. In this study, our attention to such obstructions was minimal.

Ergonomic studies in laparoscopic surgery have focused on understanding and improving the ergonomic risks and benefits involved in the primary surgeon's performances [21]. In a recent literature review, these evidence-based experimental ergonomic studies were discussed in detail, including examination of their methodological infrastructures (e.g., tasks, models, measurement systems) [13]. Additionally, the obtained data and specific applications covered in these ergonomic studies were summarized. In brief, the review investigated studies on the effects that operating room components such as display systems [22–24], instrument handle designs [25–28], operating table heights [29, 30], and instrument, scope, and task locations/

orientations [31–33] have on task performance in addition to the physical workload of primary surgeons.

The studies reviewed also quantitatively assessed levels of primary laparoscopic surgeons' skills, dexterity, and motion efficiency [27, 34, 35]. The majority of models and tasks used in the reviewed studies were designed to simulate the instrument maneuvering exercised and the surgical tasks performed by primary surgeons [29, 30, 36–39].

This review examining the current body of ergonomic research related to laparoscopic surgery showed the need to extend that analysis to support staff who have an active role in operating room procedures. The few existing studies on the task performance of camera assistants had goals quite different from ergonomic risk investigation. One study investigated the application of an electromagnetic motion tracking system for simple measurements of the movements of the camera and holding arm [40]. Another study compared motion efficiencies measured from human camera drivers and robot-assisted camera control [41].

In a series of studies focused on products rather than task performance, Van Veelen et al. [24] investigated the relation of display locations to assistants' neck movements and muscle activation during camera holding, although not comparatively [29]. In an observational survey study, Van Veelen made the statement that positioning of the camera may be associated with physical discomfort of the back [42].

The strength of the conclusions we reached about the hypotheses proposed in this study compel us to analyze further our data collection with the purpose of understanding the association of specific muscle groups to fatigue. We also have begun to develop a method for calculating the compressive joint reaction forces related to physical stress. An understanding of underlying stress and fatigue mechanisms, coupled with our current findings from WLR analysis, promises to allow for further identification and minimization of evident ergonomic risks in the execution of camera-pointing and retraction tasks performed by laparoscopic assistants.

Acknowledgments This study was supported by a grant from the U.S. Army Medical Research and Materiel Command (USAMRMC), and equipment was provided in part by U.S. Surgical. The authors acknowledge the thoughtful and careful assistance of Rosemary Klein in the editing of this article.

References

1. Bhatnager V, Drury CG, Schiro SG (1985) Posture, postural discomfort, and performance. Hum Factors 27:189–199
2. Liao MH, Drury CG (2000) Posture, discomfort, and performance in a VDT task. Ergonomics 43:345–359
3. Van Wely P (1970) Design and disease. Appl Ergon 1:262–269
4. Keyserling RM (1986) A computer-aided system to evaluate postural stress in the workplace. Am Ind Hyg Assoc J 47:641–649
5. Winter DA (1995) Human balance and posture control during standing and walking. J Biomech 3:193–214
6. Gage WH, Winter DA, Frank JS, Adkin AL (2004) Kinematic and kinemtic validity of the inverted pendulum model in quiet standing. Gait Posture 19:124–132
7. Corlett EN, Bishop RP (1976) A technique for assessing postural discomfort. Ergonomics 19:175–182
8. Kant IJ, de Jong LC, Van Rijssen-Moll M, Borm PJ (1992) A survey of static and dynamic work postures of operating room staff. Int Arch Occup Environ Health 63:423–428
9. Patkin M, Isabel L (1995) Ergonomics, engineering, and surgery of endosurgical dissection. J R Coll Surg Edinb 40:120–132
10. Reyes DAG, Tang B, Cuschieri A (2006) Minimal access surgery (MAS)-related surgeon morbidity syndromes. Surg Endosc 20:1–13
11. Berguer R, Forkey DL, Smith WD (1999) Ergonomic problems associated with laparoscopic surgery. Surg Endosc 13:466–468
12. Carswell CM, Duncan C, Seales WB (2005) Assessing mental workload during laparoscopic surgery. Surg Innov 12:80–90
13. Lee G, Lee T, Dexter D, Klein R, Park AE (2007) Methodological infrastructure in surgical ergonomics: a review of tasks, models, and measurement systems. Surg Innov 14:153–167
14. Person JG, Hodgson AJ, Nagy AG (2001) Automated high-frequency posture sampling for ergonomic assessment of laparoscopic surgery. Surg Endosc 15:997–1003
15. Berguer R, Rab GT, Abu-Ghaida H, Alarcon A, Chung J (1997) A comparison of surgeons' posture during laparoscopic and open surgical procedures. Surg Endosc 11:139–142
16. Gillette JC, Quick NE, Adrales GL, Shapiro R, Park AE (2003) Changes in posture mechanics associated with different types of minimally invasive surgical training exercises. Surg Endosc 17:259–263
17. Lee G, Kavic SM, George IM, Park AE (2006) Correlation between postural stability and performance time during fundamentals of laparoscopic surgery (FLS) tasks. Br J Surg 93(Suppl):S206
18. Savoie S, Tanguay S, Centomo H, Beauchamp G, Anidjar M, Prince F (2007) Postural control during laparoscopic surgical tasks. Am J Surg 193:498–501
19. Lee G, Park AE (2007) Development of a more robust tool for postural stability analysis of laparoscopic surgeons. Surg Endosc 22:1087–1092
20. Lee G, Kavic SM, George IM, Park AE (2007) Postural instability does not necessarily correlate to poor performance: case in point. Surg Endosc 21:471–474
21. Van Veelen MA, Jakimowicz JJ, Kazemier G (2004) Improved physical ergonomics of laparscopic surgery. Min Invas Ther Allied Technol 13:161–166
22. Smith WD, Berguer R, Nguyen NT (2005) Monitor height affects surgeons' stress level and performance on minimally invasive surgery tasks. Stud Health Technol Inform 111:498–501
23. Huber JW, Taffinder N, Russell RCG, Darzi A (2003) The effects of different viewing conditions on performance in simulated minimal access surgery. Ergonomics 46:999–1016
24. Van Veelen MA, Jakimowicz JJ, Goossens RHM, Meijer DW, Bussmann JBJ (2002) Evaluation of the usability of two types of image display systems during laparoscopy. Surg Endosc 16:674–678
25. Emam TA, Frank TG, Hanna GB, Stockham G, Cuschieri A (1999) Rocker handle for endoscopic needle drivers. Surg Endosc 13:658–661
26. Van Veelen MA, Meijer DW, Uijttewaal I, Goossens RHM, Snijder CJ, Kazemier G (2003) Improvement of the laparoscopic

needle holder based on new ergonomic guidelines. Surg Endosc 17:699–703
27. Uchal M, Brogger J, Rukas R, Karlsen B, Bergamaschi R (2002) In-line vs. pistol-grip handles in a laparoscopic simulator: a randomized controlled crossover trial. Surg Endosc 16:1771–1773
28. Emam TA, Frank TG, Hanna GB, Cuschieri A (2001) Influence of handle design on the surgeon's upper limb movements, muscle recruitement, and fatigue during endoscopic suturing. Surg Endosc 15:667–672
29. Van Veelen MA, Kazemier G, Koopman J, Goossens RHM, Meijer DW (2002) Assessment of the ergonomically optimal operating surface height for laparoscopic Surgery. J Laparoendosc Adv Surg Tech 12:47–52
30. Berguer R, Smith WD, Davis S (2002) An ergonomic study of the optimum operating table height for laparoscopic surgery. Surg Endosc 16:416–421
31. Emam TA, Hanna GB, Kimber C, Dunkley P, Cuschieri A (2000) Effect of intracorporeal–extracorporeal instrument length ratio on endoscopic task performance and surgeon movements. Arch Surg 135:62–65
32. Emam TA, Hanna G, Cuschieri A (2002) Comparison of orthodox vs off-optical axis endoscopic manipulations. Surg Endosc 16:401–405
33. Emam TA, Hanna G, Cuschieri A (2002) Ergonomic principles of task alignment, visual display, and direction of execution of laparoscopic bowel suturing. Surg Endosc 16:267–271
34. Emam TA, Hanna GB, Kimber C, Cuschieri A (2000) Difference between experts and trainees in the motion pattern of the dominant upper limb during intracorporeal endoscopic knotting. Dig Surg 17:120–125
35. Datta V, Chang A, Mackay S, Darzi A (2002) The relationship between motion analysis and surgical technical assessment. Am J Surg 184:70–73
36. Matern U, Kuttler G, Giebmeyer C, Waller P, Faist M (2004) Ergonomic aspects of five different types of laparoscopic instrument handles under dynamic conditions with respect to specific laparoscopic tasks: an electromyographic-based study. Surg Endosc 18:1231–1241
37. Hanna GB, Drew T, Clinch P, Hunter B, Shimi S, Dunkley MP, Cuschieri A (1996) A microprocessor-controlled psychomotor tester for minimal access surgery. Surg Endosc 10:965–969
38. Quick NE, Gillette JC, Shapiro R, Adrales GL, Gerlach D, Park AE (2003) The effect of using laparoscopic instruments on muscle activation patterns during minimally invasive surgical training procedures. Surg Endosc 17:462–465
39. Crosthwaite G, Chung T, Dunkley P, Shimi S, Cuschieri A (1995) Comparison of direct vision and electronic two- and three-dimensional display systems on surgical task efficiency in endoscopic surgery. Br J Surg 82:849–851
40. Riener R, Reiter S, Rasmus M, Wetzei D, Feussner H (2003) Acquisition of arm and instrument movements during laparoscopic interventions. Min Invas Ther Allied Technol 12:235–240
41. Kondraske GV, Hamilton EC, Scott DJ, Fischer CA, Tesfay ST, Taneja R, Brown RJ, Jones DB (2002) Surgeon workload and motion efficiency with robot and human laparoscopic camera control. Surg Endosc 16:1523–1527
42. Van Veelen MA, Nederlof EA, Goossens RH, Schot CJ, Jakimowicz JJ (2003) Ergonomic problems encountered by the medical team related to products used for minimally invasive surgery. Surg Endosc 17:1077–1081